"1. Study Title : ""Journey into SPACE: Evidence-based Design of an App to Reduce Digital Addiction""
Rater 1 Rating = 3 (Good)
AAR Justification: Both the REF submission and the ChatGPT version address the critical role of the SPACE app in reducing digital addiction effectively, highlighting its features like notification management and social norms feedback. However, some minor discrepancies in emphasizing specific user statistics were noted.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with the outcomes described in the REF-submitted version. Both versions mention the app's development, its behavioral change strategies, and the reduction of digital addiction (DA) through features like reducing notifications and managing FOMO. Both the AI version and the REF submission emphasize global adoption and usage of the app, highlighting the success of SPACE. However, there are slight variations in focus, such as more emphasis on interdisciplinary research in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact aligns well with most outcomes and impacts specified in the REF 2021 submission, particularly around social and technological impacts, such as addressing digital addiction (DA) through behavior-driven app design. The AI version captures the main impacts in public awareness and user engagement, along with improvements in user productivity and wellbeing, accurately reflecting outcomes stated in the original REF submission. However, specific policy implications noted in the REF version, like regulations and long-term DA management recommendations, were only partially captured by ChatGPT.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts aligned closely with the goals and evidence base described in the original REF case study, particularly regarding behavioral change strategies to counter digital addiction (DA) and design features tailored for user self-regulation. The AI content also accurately reflected the focus on reducing smartphone usage to improve well-being and productivity. However, while AI addressed key aspects such as FOMO reduction and social comparison, some finer details—like user interaction patterns over time—were more explicitly covered in the REF submission.

"
"2. Study Title : Using Macroprudential Policies to Reduce the Risk of Financial Crises
Rater 1 Rating = 3 (Good)
AAR Justification: The AI-generated research impact generally aligns with the major outcomes and impacts described in the REF 2021 submission. For instance, both versions discuss the development of macroprudential tools and the role of NiGEM in influencing global policy frameworks. However, the AI version is more generalized and misses some nuanced details regarding specific institutional impacts.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with most of the key outcomes mentioned in the REF submission, such as the influence on global financial institutions, the adoption of tools by the IMF, Norges Bank, and NIESR. However, some finer details of certain institutional impacts and specific financial models were not fully mirrored in the AI-generated version.
Rater 3 Rating = 4: Very Good
AAR Justification: Matched in Both Versions: The ChatGPT version correctly identifies key impacts of macroprudential policies, such as reducing systemic risk and influencing policy and economic stability, closely mirroring the REF's emphasis on practical, institution-level impact. REF Only: REF 2021 uniquely details specific applications and outcomes from financial institutions (e.g., NIESR, Norges Bank, IMF), demonstrating its applied, contextual accuracy. ChatGPT Only: While broader, the ChatGPT version aligns well with the objectives by projecting a range of global applications and forward-looking financial impact, though lacking specific institutional contexts found in REF.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated version accurately captures significant outcomes and effects from the original REF case, including policy influence on institutions like the IMF, OECD, and Norges Bank. This alignment reflects an understanding of the study's impact on economic policy and financial risk reduction. However, there are some specific tools and datasets mentioned in the REF submission that the AI version only broadly references, potentially missing nuanced details in the methodological alignment.

"
"3. Study Title : Utilising the human-canine relationship to support vulnerable people in the criminal justice system
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts aligned well with the outcomes described in the REF

submission. Key elements, such as the introduction of facility dogs to support vulnerable victims in courtrooms and their influence on policies like ABE interviews, were present in both versions. ChatGPT accurately identified the impact on court processes and policy changes in England and abroad. However, some details, such as the specific number of courts that adopted facility dogs and the role of international organizations, were more explicitly mentioned in the REF submission.

Rater 2 Rating = 4

AAR Justification: The AI-generated impacts accurately reflect many key aspects of the REF-submitted case study, including the use of therapy dogs to support vulnerable victims in court settings and the subsequent national and international adoption of these practices. The AI version mentions the emotional and psychological benefits, specifically addressing the implementation in the UK and international contexts. However, the AI version occasionally lacks the specific details related to the timeline and exact court systems involved, leading to a slightly less comprehensive understanding of the impact.

Rater 3 Rating = 4 (Very Good)

AAR Justification: The ChatGPT-generated impacts align well with key aspects of the REF 2021 submission, covering major domains such as the support of vulnerable populations, policy influence, and training. Both versions effectively reflect changes in court practices and emotional support outcomes from therapy dogs in the legal context. However, slight discrepancies exist in detailing the influence on international practices.

Rater 4 Rating = 4: Very Good

AAR Justification: The ChatGPT-generated impact narrative closely aligns with the key outcomes of the REF 2021 submission, accurately reflecting impacts on policy and practice within the criminal justice system. This includes mentions of therapy dogs' role in supporting victims and witnesses, improvements in the ABE interview process, and policy endorsements. The AI-generated content reflects major themes of stress reduction, improved witness comfort, and increased adoption across multiple UK and international court systems. However, some nuanced aspects, like the role of specific organizations or the detailed policy pathway, were not as fully explored in the AI version.

"

"4. Study Title : Embedding responsible practices into business by taking inspiration from the Quakers

Rater 1 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts aligned well with the key outcomes described in the REF-submitted version, especially regarding the application of Quaker Discernment (QD) in governance, leadership, and decision-making. Both versions mentioned how QD principles were embedded into governance structures of various organizations, resulting in non-hierarchical decision-making and improved team cohesion. However, there were a few nuanced differences in the phrasing and emphasis on outcomes.

Rater 2 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts closely align with the REF-submitted version in key areas such as the implementation of Quaker Business Methods (QBM) in corporate governance, including its non-hierarchical decision-making processes. Both versions highlight how QBM influences team cohesion and governance structure in secular organizations. However, minor variations were observed regarding the specific examples cited.

Rater 3 Rating = 4 (Very Good)

AAR Justification: The AI-generated content aligns closely with the original REF submission in capturing the integration of Quaker principles into business practices, specifically the adoption of Quaker Business Method (QBM) and its impacts on decision-making structures across various sectors. The AI version successfully mirrors details about stakeholder collaboration, ethical implications, and organizational culture benefits.

Rater 4 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts align well with the core impacts outlined in the REF submission, capturing key themes such as the introduction of QBM to secular organizations and its impact on governance structures, stakeholder collaboration, and ethical decision-making. Specific impacts, like improvements in organizational efficiency and stakeholder engagement, are mirrored in both versions, demonstrating a high level of alignment.

"

"5. Study Title : Hope 4 The Community CIC: Improving Lives of People Living with Long-Term Health Conditions by providing Self-Management Tools

Rater 1 Rating = 4 (Very Good)

AAR Justification: The AI-generated content aligns well with the core themes and outcomes presented in the REF submission. It mentions key components such as the impact of self-management on health outcomes, including mental well-being and the reduction of anxiety and

depression. Both reports mention peer-delivered interventions and the role of positive psychology in managing long-term conditions. However, some specifics in the original report, such as the influence on national and international policies and the partnerships with organizations like Macmillan Cancer Support, were less emphasized in the AI version.
Rater 2 Rating = 4: Very Good
AAR Justification: Both the AI-generated version and the REF submission describe the Hope Programme's impact on long-term conditions, with significant details on health, psychological, and social well-being. The REF submission emphasizes studies related to multiple sclerosis, cancer, and mental health, while the AI-generated version aligns with these key themes but sometimes lacks granular details such as the specific methodologies used in clinical trials. The alignment is strong, though the AI version occasionally generalizes certain outcomes without the nuanced context provided by the REF submission.
Rater 3 Rating = 4
AAR Justification: The AI-generated version closely aligns with the REF submission, covering core themes like patient activation, self-management, and psychosocial outcomes. Both versions highlight similar goals, including improving mental health and expanding self-management capabilities. However, AI content includes extra interpretations in policy impact and patient activation detail, which occasionally diverge from REF's core focus on self-management tools.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated content accurately reflects the main impacts in the REF 2021 submission, such as patient activation, mental well-being improvement, and self-management. It includes key aspects like support through digital platforms during COVID-19, peer-delivered interventions, and the establishment of a social enterprise. However, some specific stakeholder partnerships mentioned in the REF, like Macmillan Cancer Support, were not included in the AI summary.

"
"6. Study Title : Localisation of Electricity Generation and Use
Rater 1 Rating = 4
AAR Justification: The AI-generated content aligns well with the core outcomes and impacts described in the REF submission. It accurately reflects the policy influence of the research, particularly the role of ""Energy Local"" and its impact on community energy models and regulatory change. However, some finer details regarding specific regulatory achievements, such as the ""complex site"" loophole used in Bethesda, are not as clearly highlighted in the AI version.
Rater 2 Rating = 4 - Very Good
AAR Justification: Both the REF 2021 submission and ChatGPT's version reflect the core impacts such as the establishment of Energy Local and Exergy Devices and their influence on consumer energy savings, smart heating systems, and policy. The AI content aligns well with these themes, capturing technological and policy impacts effectively.
Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated research impact aligns closely with the REF submission, capturing the core achievements in energy efficiency, economic impact, and technological advancements. The ChatGPT version accurately reflects the project's contributions to local energy generation and the benefits for consumers and community energy efforts.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the REF-submitted impacts in key areas such as environmental sustainability, public engagement, and technological development. Specific impacts on public policy, electricity tariffs, and energy localization models were accurately captured. However, nuances in regulatory discussions and exact policy stakeholders were less developed.

"
"7. Study Title : Critical Connections pedagogical model based on multilingualism and digital storytelling boosts language learning and digital skills
Rater 1 Rating = 3: Good
AAR Justification: The AI-generated version largely aligns with the key outcomes from the REF submission, such as improving digital and pedagogical skills for teachers and enhancing students' pride in multilingualism. However, some of the more nuanced impacts, such as the specific partnerships with Cyprus and the direct feedback from international stakeholders (ECML, NRCSE), are not as fully detailed per the AI-generated version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated version closely aligns with key impacts mentioned in the REF submission, such as the development of new pedagogic skills and the creation of a global digital storytelling network. However, some specific details, such as the partnerships with Cyprus and Taiwan or the British Film Institute's role, were underrepresented in the AI version.

Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated content aligns with the original REF 2021 impacts, capturing the core themes of multilingual pedagogy, student and teacher skill development, and community engagement. Both versions recognize the significance of new pedagogical models and enhanced digital storytelling, although the REF submission provides additional context-specific details, such as the inclusion of the European Centre for Modern Languages (ECML) and the strategic partnerships with complementary schools.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts largely align with the REF 2021 submission, accurately reflecting key aspects such as pedagogical skill development for teachers, enhanced engagement in multilingual education, and the creation of a supportive digital storytelling community. Specific impacts like fostering pride in multilingualism among students and establishing international collaborations are represented in both versions, indicating strong alignment.

"
"8. Study Title : Helping to Sustain the UK's Independent Film Industry Through an Improved Risk Management Strategy
Rater 1 Rating = 3 (Good)
AAR Justification: Both the ChatGPT-generated version and the REF-submitted version addressed the central themes of risk management in the UK film industry, highlighting economic, policy, and industry impacts. However, the AI-generated content lacked specific stakeholder mentions and some key details from the REF version, such as the exact impact on institutions like the BFI and DCMS.
Rater 2 Rating = 4: Very Good
AAR Justification: Both versions, the AI-generated content and the REF submission, align closely in terms of identifying key economic, social, and policy impacts, particularly around financial risk management and support for the UK independent film sector. The AI-generated version captures core themes such as improving financial decision-making, cultural preservation, and policy influence, which are crucial components of the REF submission. However, the REF submission provides slightly more depth in some specific areas, such as stakeholder engagement strategies.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT version aligns well with key themes in the REF submission, covering significant aspects such as risk management improvements and industry-wide benefits. However, some nuanced details, such as specific references to legislative changes influenced by the research, are more extensively detailed in the REF submission.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with many key outcomes and impacts outlined in the original REF 2021 case study. Both the AI and REF versions cover specific elements such as policy influence, industry impact, and strategic recommendations on risk and data management, highlighting the value of blockchain for IP registration and industry data transparency. However, certain detailed mechanisms of change, especially those related to the political context, are more fully captured in the REF submission.

"
"9. Study Title : Changing practice and improving wellbeing through immersive vocal art
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated research impacts generally align well with the REF-submitted version. Core elements such as enhancing creative vocal techniques for children and disabled groups, the influence on cultural practices, and contributions to queer sound studies are accurately reflected. However, a few nuances related to stakeholder engagement and practical applications are missing from the AI version.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with the REF submission, especially in highlighting the cultural, social, and academic impacts of the research. Both versions accurately reflect the contribution to cultural practices and audience participation through voice-expanding technologies. However, some nuanced details, like the specific technological and artistic applications (e.g., the Resonant Tails impact on disabled children), are more deeply elaborated in the REF version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align closely with the core objectives outlined in the REF submission, notably around improving wellbeing through innovative vocal practices and immersive technologies. It accurately reflects the project's focus on enhancing social and cultural engagement via non-normative sound explorations, though it misses specific references to nuanced elements such as ""voice-expanding mirror"" technology for PMLD children and certain

institutional partnerships.
Rater 4 Rating = 4
AAR Justification: The ChatGPT-generated impacts reflect key project elements, such as the creation of immersive vocal performances to engage children and marginalized groups, closely aligning with the REF submission's descriptions of ""Your Vivacious Voice"" and ""Resonant Tails."" Specific mentions of enhanced participation, quality of life improvements, and use of interactive technologies, such as the VT and app-based voice activities, align well with the AI-generated analysis. However, the AI content occasionally generalized terminology around inclusivity and creative engagement without specifying nuances, such as the impact on disabled youth or the distinct role of installations in Science Centres that were highlighted in the REF version.

"
"10. Study Title : Safewards: Increasing Safety on Psychiatric Inpatient Wards
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impacts align well with the key themes of the REF-submitted version. The ChatGPT version captured essential aspects like the reduction of conflict and containment in psychiatric settings, the impact on staff-patient relationships, and contributions to health policies, particularly noting national recommendations such as NICE. However, there are some nuanced details missing or less accurately conveyed, which affects the rating slightly.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligned closely with the core impacts described in the REF submission. The AI accurately reflected key outcomes such as reductions in conflict and containment, improvements in staff safety, and the broad implementation of the Safewards model across the UK and internationally. Both versions highlighted the model's inclusion in the NICE guidelines and the positive endorsement by governmental bodies like NHS England and the Department of Health.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated content generally aligns well with the REF 2021 submission, capturing core aspects such as the focus on safety in psychiatric settings and the reduction of conflict and containment. Specific impacts on policy (e.g., NICE guidelines) are acknowledged, though with less detail compared to the original, particularly in how Safewards' integration extends to varied institutional policies across the UK and international systems.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts accurately reflect the REF 2021 study's focus on reducing conflict and containment in psychiatric settings. It effectively references major outcomes like policy influence, patient-staff interactions, and ethical considerations. It also matches REF's emphasis on the NICE guideline recommendation for Safewards and mentions similar policies in international contexts, showing strong alignment. However, the AI version lacks some specific instances of implementation, such as Safewards' widespread use in individual UK trusts and its influence on specific NHS frameworks.

"
"11. Study Title : Contemporary documentary practices: historical perspective and interdisciplinary approaches – the International Research Centre for Interactive Storytelling (IRIS)
Rater 1 Rating = 4
AAR Justification: The AI-generated research impacts are largely aligned with the outcomes and impacts described in the REF submission. Both versions recognize the influence of the research on academic fields such as interactive documentary practices, transdisciplinarity, and social impact through civic engagement. However, while the AI version does capture most of the important outcomes, it occasionally lacks the depth of specific examples provided in the REF submission, such as detailed module development at Leeds Trinity and other institutions.
Rater 2 Rating = 4 – Very Good
AAR Justification: Both the REF-submitted version and the AI-generated content reflect the interdisciplinary and socially-engaged nature of the research, emphasizing interactive documentary methods and their societal impacts, such as community empowerment and digital archival of marginalized voices. However, some nuances in the depth of community engagement and archival methodologies were more explicitly detailed in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts largely align with the REF 2021 submission, accurately representing key research impacts across domains such as social, cultural, and policy impact. While a few nuanced aspects of project impacts, like specific module influences or certain community engagements, were underemphasized in the AI version, core themes of transdisciplinary influence and interactive documentary preservation were well-aligned.

Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impact aligns closely with the REF submission's key themes, including academic transdisciplinarity and social impacts through digital archiving and community engagement. While the AI captures most core outcomes, it occasionally generalizes impacts that the REF details explicitly, such as specific modules impacted within Leeds Trinity and other institutions.

"
"12. Study Title : Improving patient outcomes through better project management of clinical trials
Rater 1 Rating = 4
AAR Justification: The AI-generated content accurately reflects many of the impacts described in the REF 2021 submission, particularly around improvements in clinical trial management, cost savings, and stakeholder relationships between pharma companies and CROs. For example, both versions mention H. Lundbeck A/S using the CURED framework to improve their project management KPIs. However, there are instances where the ChatGPT version lacks specific examples or fails to go into as much detail as the REF version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impact largely aligns with the original REF submission. Both versions discuss the significant impact of the CURED framework on project management within clinical trials. The AI content reflects well on key aspects like cost savings and time efficiency, which were central to the original submission. However, some nuances related to stakeholder-specific insights, particularly from Lundbeck and GSK, were not fully captured in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the REF 2021 submission across most key themes, accurately capturing impacts on drug development efficiency, project management, and the economic implications of improved clinical trial processes. Both versions emphasize cost savings, time efficiencies, and the new methodologies for managing clinical trials. However, the AI version generally did not specify certain organizational names or in-depth process details (e.g., LEO Pharma's specific KPIs) included in the REF submission.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts are well-aligned with many themes in the REF submission, especially the emphasis on efficient project management in clinical trials. Both versions acknowledge key stakeholders, including pharmaceutical companies and CROs, and address the importance of the client-supplier relationship. The AI analysis effectively reflects these core themes but lacks specific company instances cited in the REF (e.g., H. Lundbeck's re-evaluation of metrics and GSK's global training rollout). Including these would enhance the alignment and strengthen the evidence base.

"
"13. Study Title : Ensuring the Fair Treatment of Open Banking Customers
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated research impacts align closely with the original research outcomes, particularly in areas such as consumer consent, privacy, and dynamic consent mechanisms in Open Banking. The AI version accurately captured key impacts relating to consumer protection, the use of APIs, and the ethical considerations associated with data sharing.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated content closely aligns with the REF-submitted version, capturing key themes like consumer data protection, dynamic consent, and regulatory impact on Open Banking. Both versions emphasize the significance of consent management and consumer protection in financial data sharing.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with the primary outcomes of the REF submission, especially in focusing on consumer data protection and privacy within Open Banking. Key points like ""dynamic consent"" and ""informed consent as a basis for consumer protection"" are mirrored accurately in the AI version. However, nuanced insights such as regulatory changes instigated by the FCA and FSCP through targeted interventions, are detailed more in the REF version.
Rater 4 Rating = 4
AAR Justification: The AI-generated content aligns closely with the key impacts described in the REF case, particularly around influencing FCA policy, consumer protection, and Open Banking consent practices. It identifies the importance of informed consent and references consumer and privacy advocacy, consistent with the REF. However, minor variances exist, where specific references to regulatory bodies or precise dates could enhance exact alignment.

"

"14. Study Title : Advancing Movement Practices in Doctoral and Professional Contexts
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated version aligns well with the research impacts mentioned in the REF submission. It covers the core aspects of the impact in advancing knowledge in practice-as-research (PaR), particularly in creative arts pedagogy and professional contexts. The AI report accurately reflects impacts on academic, social, and policy domains. However, there are some nuances in specific case studies that are more detailed in the REF version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impact reflects key elements of the REF-submitted impact. It successfully identifies major impacts on creative practices, higher education, and doctoral research environments. However, there are minor discrepancies in some details. For instance, while the AI captured the overarching goals and the CAP's influence on artistic research and education, it slightly underrepresented the depth of international engagement mentioned in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with key aspects from the REF-submitted version. The core concepts, such as the impact on creative practice and the advancement of practice-as-research (PaR) through the CAP model, are accurately captured. However, some nuanced contributions in terms of international partnerships and specific organizational impacts (e.g., Dance4 and ADiE) are only broadly referenced in the AI version.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the REF 2021 submission in accurately identifying key themes and impacts. It accurately highlights CAP's influence on dance pedagogy, creative practice, and HEI curricula integration, reflecting Midgelow's role in advancing the dance field through CAP, especially in non-academic applications.

"

"15. Study Title : Digital Twin Specification, Design and Application
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the research outcomes described in the REF submission, particularly in technological advancements, such as the LEAP framework and its application in enterprise modeling. Both versions recognize the collaboration with TCS, the development of new enterprise tools, and the impact on TCS's operations and market strategies.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with the overall themes and outcomes mentioned in the REF submission. Both versions highlight critical industry collaborations, especially with TCS, and advancements in digital twin technology. The AI-generated version correctly reflects the development of the digital twin for enterprise modeling, agent-based systems, and structural health monitoring.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align well with the original REF submission, capturing core themes in both Digital Twin (DT) applications for structural health monitoring and enterprise simulation modeling. The AI content reflects significant collaborations and the development of the TwinX™ product, capturing impacts on organizational modeling and enterprise optimization. However, while well-aligned, some specificity in real-world applications (e.g., COVID-19 modeling in Pune) lacks detail in the AI content.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts largely align with the research impacts presented in the REF submission, especially in terms of enterprise modeling and technology transfer between academia and industry. The AI-generated impacts correctly reflect key areas like the development of digital twin technology and its application in enterprise settings. However, while the AI version captures the overarching themes, some nuances specific to the collaborations with TCS and the application of ESL technology are less detailed.

"

"16. Study Title : Being in Touch: Inspiring Cultural Engagement through Creative-Critical Writing
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated content closely aligns with the outcomes described in the REF 2021 submission, especially in capturing the academic and cultural impacts. Both highlight the engagement with cultural participation through creative-critical writing and the academic contributions of tactile poetics. However, there are nuanced differences in certain specifics, such as detailed examples of BBC collaborations and community workshops.
Rater 2 Rating = 4 (Very Good)

AAR Justification: The ChatGPT-generated content aligns well with the key themes of the REF submission, including cultural engagement and creative-critical writing. For instance, both the REF and ChatGPT versions highlight the BBC collaborations and the impact on public discourse. However, some specific details about the types of audiences and exact collaborations, such as with the BT Archives, are more detailed in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns closely with the key aspects highlighted in the REF submission, particularly in how it presents cultural engagement, creative programming, and interdisciplinary research impact. Both versions cover domains such as media impact, public engagement, and creative programming, with specific events and workshops reflecting public engagement. However, the AI-generated impact sometimes lacks the precise examples and extent of media influence presented in the REF submission.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align with the REF-submitted version well, especially in capturing the broad themes of cultural engagement and innovation in creative writing. Key impacts like media influence (BBC programming) and public engagement through workshops match both versions. However, the REF submission provides more specificity in audience impact (e.g., number of listeners, qualitative participant feedback). This detailed specificity in the REF document illustrates direct public engagement more precisely, whereas the AI version summarizes impact broadly.

"
"17. Study Title : Shaping crime prevention policy and strategy to sustain the crime drop and reduce domestic burglary
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the outcomes and impacts described in the REF submission. Both versions discuss the role of security in reducing crime and burglary, particularly focusing on the ""security hypothesis"" as a core explanation for crime reduction. However, some nuanced differences exist in how specific impacts on policy and community initiatives are presented. The AI-generated version provides accurate but slightly broader interpretations of impacts, whereas the REF submission is more detailed in citing specific events and government actions.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact statements accurately reflect many of the core impacts found in the REF submission, such as shaping national and regional policies on crime reduction, particularly in areas like burglary prevention and influencing crime reduction strategies. Both versions emphasize the ""security hypothesis"" as a key driver for the crime drop. However, there were slight differences in the articulation of specific stakeholders involved, such as detailed contributions to the Modern Crime Prevention Strategy.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the original studies' findings and goals, correctly reflecting key impacts on policy and strategy without significant deviation.
Rater 4 Rating = [4] Very Good
AAR Justification: The ChatGPT-generated impacts align closely with the outcomes described in the REF 2021 submission, particularly regarding the influences on national crime policy and local crime reduction initiatives. Specific aspects, like policy adjustments based on security hypothesis findings, match in both versions, demonstrating a consistent alignment with intended project outcomes.

"
"18. Study Title : Building local socio-economic impacts into the assessment of major energy projects
Rater 1 Rating = 4: Very Good
AAR Justification: The ChatGPT-generated impacts aligned well with the major themes in the REF submission, particularly around the socio-economic impacts of major energy projects on local communities, employment, housing, and services. However, the REF submission provided more specific details, such as stakeholder responses (e.g., Somerset County Council's direct feedback), which were not as thoroughly reflected in the ChatGPT version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the REF submission in key areas, such as socio-economic impacts of large energy projects like Hinkley Point C and Sizewell B. Both versions emphasize impacts on local employment, housing, and public services.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated content closely aligns with the original impacts described by the research team, capturing key elements like socio-economic impacts on local communities and policy implications.

Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts generally align well with the REF-submitted content, capturing core aspects of socio-economic and environmental impacts associated with large energy projects. Specific mentions of stakeholders like local governments and industries, as well as methodological innovations, demonstrate strong alignment. However, nuances such as specific numerical data on project timelines and cost allocations from the REF version were less emphasized in ChatGPT's content.

"

"19. Study Title : Strengthening global and national policies on performance-based and innovative health financing in low-income and fragile settings
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the core themes and objectives of the REF 2021 submission. Both versions focus on the impacts of performance-based financing (PBF) on healthcare systems, particularly in low- and middle-income countries (LMICs). Key areas such as policy influence, stakeholder engagement (e.g., WHO, World Bank), and the practical application of PBF models are present in both versions. However, some nuances around specific organizational collaborations (e.g., Global Fund to fight HIV/AIDS) are more detailed in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the overall research themes of performance-based financing (PBF), particularly its impact on health financing policy and practice at both global and national levels. Both versions focus on major international organizations like the WHO and World Bank, and on PBF implementation in fragile settings. However, some specific details, such as particular case studies on countries like Zimbabwe and Sierra Leone, were less developed in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the specific areas of influence highlighted in the REF 2021 submission, particularly in domains like health policy impact, international collaboration, and national-level application. Both versions consistently mention impacts on global health organizations (e.g., WHO, World Bank), the healthcare policy domain, and targeted health financing. ChatGPT-generated content has accurately covered the foundational aspects and aligns with key policy and practitioner-level impacts. However, some nuanced details in direct collaborations were less emphasized in the AI version.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the REF-submitted case study's primary themes, accurately reflecting impacts on health financing, policy development, and international collaboration. Both versions emphasize collaboration with global organizations (e.g., WHO, World Bank) and policy influence in low-income settings. However, the AI analysis slightly generalized specific impacts on the World Bank's substantial investment adjustments and the Global Fund's revised strategies, which are detailed in the REF submission.

"

"20. Study Title : Transforming the Accessibility and Discoverability of Millions of Archival Television Programmes
Rater 1 Rating = 4 (Very Good)
AAR Justification: Both versions align well in covering the major themes of archival TV restoration and usability, particularly around the impact of Box of Broadcasts and EUscreen as essential tools in education and research. The AI-generated content successfully captures most of the significant aspects mentioned in the REF submission, especially regarding the preservation and public accessibility of television archives, educational uses, and technological contributions to media heritage. However, some nuanced details, such as the precise scope of public engagement and the role of historic technologies, are less emphasized in the AI version.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with the core themes of the original research case study. It successfully captures key impacts such as the preservation and restoration of archival TV content and the development of access platforms like Box of Broadcasts and EUscreen. The AI version also highlights the societal impact, including public accessibility and educational uses, though it occasionally lacks specific historical and technical details mentioned in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impacts align closely with key themes from the REF submission, such as advancing archival accessibility and public policy impact. AI content accurately reflects the emphasis on preservation and usability of TV archives for academic and public access. The alignment covers core objectives, like Box of Broadcasts and EUscreen, though

some nuanced aspects of technological input and metadata standardization in the REF submission were less prominent in the AI version.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts largely align with the themes in the REF 2021 submission, capturing key outcomes such as educational accessibility, cultural preservation, and enhanced usability of digital archives. Both versions emphasize the importance of accessible archival material for academia and the public, demonstrating significant congruence in overarching goals.

"
"21. Study Title : Improving quality of life for patients with Parkinson's disease
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT version accurately captures most of the key impacts mentioned in the REF submission. For example, it recognizes the improvements in clinical treatment practices, such as the use of levodopa, and its broad application across healthcare systems. Both versions mention significant healthcare cost reductions and guideline changes in the UK and Canada.
Rater 2 Rating = 4
AAR Justification: Strong alignment with key impacts described in the REF submission but lacking in detailed national guidelines and policy impact.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns closely with the REF submission by accurately referencing improvements in quality of life for PD patients through optimized treatment protocols and enhanced patient care. Both versions note significant changes in guidelines (e.g., NICE and Canadian) that affect clinical practice, though the REF version provides more detailed citations and specific examples (e.g., specific NICE recommendations). This depth of detail is partially captured in the AI version but lacks some references to exact guideline recommendations.
Rater 4 Rating = 4
AAR Justification: Both versions largely align with the project's main achievements, including changes in UK and Canadian guidelines for Parkinson's disease, the establishment of levodopa as first-line treatment, and the integration of DBS surgery for advanced stages. The AI-generated version accurately reflects these core impacts on clinical practices and guidelines, demonstrating high alignment with the key aspects. However, the original submission provides more granular details, particularly on specific NICE guidelines influenced by study results.

"
"22. Study Title : Improving Treatment for Women Suffering from Endometrial Hyperplasia
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impacts largely align with the REF submission. Both versions describe the health and well-being improvements brought about by the use of the LNG-IUS (levonorgestrel-releasing intrauterine system) over hysterectomy as the first-line treatment for endometrial hyperplasia (EH). The AI version accurately reflects the impact on clinical guidelines, treatment protocols, and patient outcomes in terms of fertility preservation, which is mentioned in the REF submission.
Rater 2 Rating = [Rating: 4]
AAR Justification: The AI-generated content aligns with many core aspects of the REF submission. It accurately reflects the primary health impacts such as the reduced need for invasive surgeries and the shift toward hormone-based treatments. It captures the policy influence on UK and international clinical guidelines related to EH treatment and fertility preservation. However, certain patient-centered details and specific statistics on treatment uptake were missing or less emphasized.
Rater 3 Rating = 4
AAR Justification: The AI-generated version aligns well with most key impacts noted in the REF submission, particularly regarding treatment improvements, healthcare access, and patient empowerment. It covers primary goals and the non-invasive treatment shift effectively. However, it occasionally generalizes specific statistics and lacks granular detail from certain REF sections (e.g., shifts in the proportion of women opting for non-surgical options over time).
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated content captures the core themes of the REF-submitted version, such as the shift from hysterectomy to hormone coil (LNG-IUS) treatments, emphasizing improved patient outcomes, especially for non-atypical endometrial hyperplasia (EH). It reflects well on the procedural shift in clinical guidelines, aligning with REF's detailed account of increased acceptance and adherence to non-invasive options among patients. However, certain subtleties in case data—for instance, patient testimonials and specific treatment adoption statistics—are less granular in the AI-generated content, impacting complete alignment.

"

"23. Study Title : Heritage of the first farmers
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the major themes and outcomes presented in the original REF submission. Both versions highlight key impacts related to heritage conservation, museum curation, community engagement, and educational benefits. For example, both discuss the role of Konya Museums, site preservation, and visitor education.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the outcomes described in the REF submission, particularly in its representation of the academic and societal impacts of the ""Boncuklu"" project. The AI content accurately captures the importance of the Konya Museums, the site's archaeological protection, and the project's influence on curatorial activities and educational resources. However, some nuanced details, such as specific testimonials from local villagers and museum directors, were not reflected in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impact aligns closely with the original REF 2021 submission, covering primary outcomes such as heritage preservation, archaeological contributions, and educational impacts. For instance, both versions acknowledge the increased public awareness of the Turkish Neolithic period through media exposure, museum exhibitions, and local educational outreach. However, the REF submission provides additional specific references to local cultural preservation efforts that were not fully captured in the ChatGPT version, which slightly limits its rating to ""Very Good.""
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts are well-aligned with the REF 2021 impact submission, particularly in areas related to public engagement, education enhancement, and contributions to archaeological preservation efforts. The alignment extends to documenting the project's influence on museum exhibitions and the preservation of heritage, reflecting accurate representation of project activities and outcomes. However, specific nuances—such as the distinct roles of local authorities and the explicit testimonies from visitors and local communities about cultural significance—were more detailed in the REF submission.

"

"24. Study Title : Mathematical modelling of an aneurysm sealing system triggers patient safety policy that withdraws surgical practice from the NHS
Rater 1 Rating = 4
AAR Justification: The ChatGPT-generated impacts closely align with the REF submission, particularly in terms of explaining the removal of the Nellix EVAS system from the NHS due to safety issues identified through mathematical modeling. Both versions correctly highlight the implications for patient safety, regulatory policy changes, and the number of affected patients.
Rater 2 Rating = 5 (Excellent)
AAR Justification: The AI-generated impacts match closely with the REF submission. The AI report correctly identifies key outcomes, such as the withdrawal of the Nellix EVAS system from NHS practice due to patient safety risks, based on mathematical modeling research. Both versions discuss the crucial impacts on clinical practice, health policy, and patient safety, maintaining alignment throughout.
Rater 3 Rating = 4
AAR Justification: Both versions align closely on key impacts related to patient safety, regulatory influence, and healthcare practices. The REF 2021 submission provides extensive details on patient outcomes, mentioning withdrawal policies and NHS practices, while the AI-generated impacts include broader health and regulatory implications without as many specific data points. AI-generated outputs correctly capture major impact areas but miss some specific NHS procedural details.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the REF 2021 case study's outcomes, specifically noting patient safety concerns and regulatory impacts. Both versions address the shift to alternative treatments and enhanced surveillance for affected patients. However, some clinical nuance is missing in the AI version, such as specific failure rates and rupture statistics associated with the Nellix EVAS device.

"

"25. Study Title : Financial and efficiency improvements from socio-technical digitalization of costing and procurement in the built environment
Rater 1 Rating = 4
AAR Justification: The AI-generated impact report aligns well with the core themes of the REF submission, such as the focus on BIM adoption, supply chain optimization, and socio-technical digitalization of costing. However, some specific examples and detailed outcomes, especially

related to policy changes and financial savings highlighted in the REF submission, are missing in the AI version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated content accurately reflects the key themes presented in the REF submission, particularly the impact on costing and procurement improvements within large construction companies like Willmott Dixon and NG Bailey. The AI also identifies key stakeholders and sectors impacted, such as the construction supply chain and governmental policy regarding procurement.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align closely with the REF submission, capturing the core themes, such as financial improvements, digital transformation in supply chain management, and the impact on costing and procurement processes. Both versions highlight the integration of digital practices and cultural shifts in construction companies, as well as stakeholder impact across supply chains. However, some nuances, particularly around policy influence and the specific actions taken by industry leaders, are less detailed in the AI-generated version.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align closely with the case study's description, capturing critical aspects like the financial savings achieved by Willmott Dixon and NG Bailey through digital transformation and cultural shifts in supply chain management. Additionally, the AI version reflects on Boyd's impact on government policy, particularly with the Procuring for Value report, though it lacks the depth seen in the REF version regarding specific stakeholder statements and in-depth financial implications.

"
"26. Study Title : Evaluating Effectiveness
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated research impacts align well with the outcomes described in the REF submission. Both versions emphasize the shift from an intervention-centered approach to a context-centered one, particularly in evidence-based policy and the broader impacts on international development, medicine, and education. However, certain nuanced aspects, like detailed stakeholder engagement pathways or specific case examples from the REF submission, are less explicit in the AI version.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated research impact aligned well with key aspects of the REF submission, particularly in describing the influence of Cartwright's work on middle-range theory (MRT) and the shift from an intervention-centered to a context-centered approach in policy-making. Both versions emphasize the impact on international development, education, and healthcare policies. However, some examples were more detailed in the REF submission, such as specific collaborations with key institutions like the World Bank and NICE.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impact aligns closely with key themes in the REF-submitted impact, notably emphasizing the limitations of RCTs and advocating for context-centered approaches. However, slight variances exist in how specific aspects of the impact pathways, particularly related to stakeholder engagement and the influence on health and policy evaluation frameworks, were represented.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT impacts align closely with the REF submission in reflecting Cartwright's focus on context-specific approaches and the limitations of RCTs in evidence-based policy (EBP). Both versions acknowledge Cartwright's influence in policy-making, particularly in sectors like international development and healthcare. However, some nuances about specific collaborations (e.g., partnership with CEDIL) and individual testimonial quotes are more detailed in the REF submission.

"
"27. Study Title : Seeing beyond the wheelchair: Pioneering education and higher aspiration promotion for boys and men with Duchenne Muscular Dystrophy
Rater 1 Rating = 4 (Very Good)
AAR Justification: ChatGPT accurately reflects key aspects such as educational and psychosocial interventions for DMD, but lacks some details on specific partnerships and policies.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impacts accurately reflect many key aspects mentioned in the REF submission, such as the focus on improving educational and psychosocial support for boys with Duchenne Muscular Dystrophy (DMD), as well as contributions to policy changes and public engagement. However, a few nuances, such as specific policy outcomes and more detailed examples from educational support (e.g., specific numbers of professionals trained, conferences held) are not as clearly detailed in the AI version.

Rater 3 Rating = 4
AAR Justification: The AI-generated content generally aligns well with the primary themes and areas of impact identified in the REF submission, including psychosocial well-being, educational needs, and transition to adulthood. Specificity on certain psychosocial and educational interventions is closely mirrored. The AI output occasionally lacks direct citation of particular quantitative measures provided in the REF.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts accurately reflect core impacts outlined in the REF submission, especially in the areas of educational and psychosocial support. The ChatGPT impacts cover multiple foundational aspects such as educational interventions, peer support, and special education services for children with DMD, aligning well with the project's core aims. Minor discrepancies are noted, however, which prevent a perfect score.

"
"28. Study Title : The Coffee Historian: Achieving Impact Through Industry Collaboration, Education, and Public Engagement
Rater 1 Rating = 4 (Very Good)
AAR Justification: Both versions accurately reflect the significant collaborations with industry partners such as Nespresso and the development of the Italian coffee culture globally. Key industry and public engagement activities described in the REF submission are also captured in the ChatGPT version, including discussions on the impact on coffee culture and business practices.
Rater 2 Rating = 4 - Very Good
AAR Justification: The ChatGPT-generated impacts align closely with the REF submission in most areas. Both versions emphasize Morris's impact on the coffee industry through collaborations (e.g., Nespresso, Gruppo Cimbali) and educational contributions to the sector. ChatGPT accurately reflects key outcomes such as product development support, educational modules, and industry influence.
Rater 3 Rating = 4 (Very Good)
AAR Justification: Both versions align well, particularly in their descriptions of cultural and educational impacts, industry collaborations, and public reach. While the ChatGPT version accurately reflects the core aspects—such as influence on coffee culture, international partnerships, and the educational modules with industry insights—it misses some nuanced impacts emphasized in the REF submission.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the REF-submitted impacts, covering key areas such as industry collaboration, educational contributions, and public engagement. For instance, both versions highlight the role of coffee history in the development of educational materials and industry practices, particularly the modules and presentations used by Nespresso. However, the AI version includes a broader reflection on global trends that aligns well but may omit some of the specific impacts cited in the REF submission, such as contributions to the Italian coffee machine heritage.

"
"29. Study Title : Discovering Ted Hughes's Yorkshire
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with many aspects of the REF submission, particularly the emphasis on the cultural and educational impacts. Both versions discuss how the research enhances understanding of Ted Hughes's life and works, particularly in Yorkshire, and how this understanding shapes stakeholders' operations and cultural experiences in the region.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts are well aligned with the REF submission. Both emphasize the development of collaborations, cultural engagement, and local heritage tourism, especially the Ted Hughes Network (THN) initiatives. However, while the AI-generated version accurately reflects key aspects such as public engagement and the impact on local authorities and communities, it lacks detailed examples provided in the REF submission, such as specific comments from local council members, stakeholders, and beneficiaries.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align closely with the REF-submitted impacts, covering main themes such as local cultural enhancement, heritage appreciation, and collaboration among local organizations. The AI effectively captured the central impact on public cultural appreciation and heritage-focused tourism.
Rater 4 Rating = 4
AAR Justification: The ChatGPT-generated impacts align well with the REF 2021 submission, capturing essential themes such as cultural engagement, knowledge dissemination, and local

economic stimulation. However, minor inconsistencies exist in reflecting the exact historical nuances and specific stakeholder interactions detailed in the REF submission.

"

"30. Study Title : Empowering Indigenous Self-Representation for the Emberá People of Panama
Rater 1 Rating = 4 (Very Good)
AAR Justification: Both the REF submission and the AI-generated version accurately address the key impacts of the research on cultural preservation, indigenous self-representation, and the use of tourism as a platform for economic and cultural empowerment. However, some nuances in the legal aspects (e.g., the land claims) are better detailed in the REF submission, while the AI version focuses more on broad cultural impacts without sufficient depth in certain cases.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated version aligns closely with the core themes presented in the REF submission. Both versions discuss the enhancement of tourism, cultural identity, and legal representation for the Emberá people. The AI version correctly reflects the core outcomes and impacts like cultural preservation and political empowerment of the Emberá. However, there were a few minor gaps regarding the specific legal precedents cited in the REF submission.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated content aligns well with key aspects of the REF submission, covering essential areas such as the use of tourism for cultural empowerment and the impact of legal documentation for ancestral land claims. Both versions address how Professor Theodossopoulos's research helped shape the Emberá's tourism practices and supported a significant legal precedent for Indigenous land rights. However, some granular details, such as the documented consultations with Panamanian authorities and the Danish NGO's role, are more explicitly covered in the REF submission.
Rater 4 Rating = [4] Very Good
AAR Justification: ChatGPT's AI-generated content aligns well with REF 2021 by highlighting the anthropological and representational focus of Professor Theodossopoulos's work with the Emberá. However, minor discrepancies exist regarding legal processes in land claims and depth in genealogical specifics used in REF. This indicates a high but not absolute accuracy.

"

"31. Study Title : Enhancing Understanding of the Foreign and Security Policy Implications of Brexit for Government and the Wider Policy Community
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated version accurately captures the overall themes and key impacts of Brexit on UK foreign and security policy, particularly in terms of its implications for policy, trade, and security relations between the UK and EU. It aligns well with the REF submission in most areas, although it misses some specific references to direct contributions in parliamentary hearings and stakeholder testimonies, which are emphasized in the REF version.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated impact aligns well with the core aspects of the REF submission. Both versions discuss Brexit's implications for UK-EU foreign, security, and defense policies. The AI version also reflects key impacts on policy-making, such as informing government bodies, think tanks, and the media, accurately reflecting the broad influence of the research. However, there are minor discrepancies, such as fewer detailed examples in the AI version.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts largely align with the REF-submitted version, addressing key themes such as Brexit's implications for UK-EU relations in foreign, security, and defense policy. The AI's focus on impacts in policy and security is accurate and resonates well with the original case study's detailed engagement with UK governmental bodies. However, the AI version lacks some specific references to Whitman's contributions to parliamentary committees and collaborations with think tanks.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact demonstrates a strong alignment with the core objectives and outputs described in the REF submission, particularly in terms of its focus on the implications for UK-EU defense collaboration and foreign policy. The AI accurately reflects the emphasis on advising UK government bodies and think tanks, though it does slightly lack the specificity found in the REF version's parliamentary engagement details.

"

"32. Study Title : Worldwide Improvements in Policing due to Increased Sales of Facial Composite Software
Rater 1 Rating = 4 (Very Good)
AAR Justification: Both the REF submission and ChatGPT version align in terms of the core

impacts described, particularly in improving police procedures globally through the use of innovative facial composite software (EFIT-V and EFIT6). They both emphasize the wide-scale adoption of the software in police constabularies across multiple countries, and improvements in policing practices. The AI-generated content reflects this impact accurately but lacks the depth of specific user testimonials and precise statistics provided in the REF submission.

Rater 2 Rating = 4

AAR Justification: The AI-generated impacts align well with the REF submission on key points such as improvements in policing practices, the widespread use of the software globally, and its commercial success. Both versions mention the enhanced ability of law enforcement to create facial composites more effectively. However, the AI version tends to be broader, sometimes lacking the specific detail found in the REF version regarding exact sales figures or police testimonials.

Rater 3 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts align closely with the primary goals of the REF-submitted case study, such as the emphasis on the software's impact on policing and criminal investigations globally. While AI content captures core aspects like advancements in recognition-based facial composites and their adoption by law enforcement, certain specifics, such as feedback from individual police constabularies, are missing.

Rater 4 Rating = 4: Very Good

AAR Justification: The AI-generated impacts align substantially with the REF 2021 submission, accurately reflecting the global impact on law enforcement and adoption metrics, as well as improvements in investigative efficiency. However, certain nuanced aspects, like specific user feedback and case examples, are less detailed in the AI-generated version.

"

"33. Study Title : Improving the Physical Wellbeing of the Police Force

Rater 1 Rating = 4 (Very Good)

AAR Justification: The ChatGPT-generated impacts closely align with the REF-submitted case study. Both versions discuss the primary goals of improving police force wellbeing, the development and success of the Fitness Mentors program, and collaboration with stakeholders like Lincolnshire Police. However, the AI-generated content sometimes lacks the specific level of detail found in the REF version.

Rater 2 Rating = 4: Very Good

AAR Justification: The AI-generated version aligns well with the core aspects of the research. Both versions focus on the Fitness Mentors Programme and its contribution to staff well-being and physical fitness, highlighting the collaboration with Lincolnshire Police. However, the REF version provides more specific examples, such as the role of key individuals like Chief Constable Skelly and detailed participant feedback. The AI version reflects these accurately but omits some interpersonal and procedural nuances.

Rater 3 Rating = 4

AAR Justification: The AI-generated version accurately aligns with the original case study's themes, particularly the impact on police workforce wellbeing and fitness initiatives. Specifics like the collaboration with Lincolnshire Police and the creation of the 'Fitness Mentors' program are well-captured. However, some nuances of the evaluation and subsequent policy integration were less detailed.

Rater 4 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts demonstrate strong alignment with the core objectives and reported outcomes from the REF submission. Both versions emphasize collaborative efforts with Lincolnshire Police to enhance police well-being through structured physical activity initiatives, including the 'Fitness Mentors' program. However, certain nuanced aspects, like the specific roles of individual contributors such as Dr. Henderson, were more elaborately detailed in the REF submission.

"

"34. Study Title : Raising Maori students' achievement in secondary schools in Aotearoa New Zealand

Rater 1 Rating = 4

AAR Justification: The AI-generated content closely matches many of the impacts described in the REF submission. Key areas such as the focus on culturally responsive pedagogy, teacher professional development, and student engagement are well captured. However, some nuanced points, particularly around specific community relationships and leadership development within the Maori context, are underrepresented.

Rater 2 Rating = 4 (Very Good)

AAR Justification: Both the REF submission and the AI-generated impacts align closely in key areas such as the importance of socio-cultural pedagogies, school reform, teacher training, and student engagement. The AI version reflects the overall impact of these initiatives with

accuracy, though it does not dive as deeply into specific numerical data or reports, which slightly reduces its accuracy in certain statistical details. For example, the AI version captures the role of culturally responsive pedagogies but lacks specific mention of the 80% of schools that developed evidence-informed action plans, which is well-documented in the REF submission.

Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT version reflects the majority of core impacts noted in the REF submission, particularly regarding educational reforms, teacher-student engagement, and culturally responsive teaching. While it aligns closely with the REF content, some specific school-level data and nuances about community relationships in KEP were more detailed in the REF version.

Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the major themes and objectives of the REF 2021 submission, capturing core themes such as Māori student engagement, curriculum reform, and teacher knowledge improvement. The AI version accurately reflects culturally responsive pedagogy and emphasizes school-community relationships. While it covers the program's overarching impact on Māori students, some specific initiatives like ""shadow coaching sessions"" and ""reflect-review-act meetings"" were only partially matched, reducing the rating to 4.

"
"35. Study Title : Improving patient outcomes and treatment guidelines through the study of Hepatitis C

Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts accurately reflect key aspects of the REF-submitted version, such as the impact on clinical practices, public health policies, and treatment guidelines for HCV. Both documents emphasize improved patient outcomes, the importance of the cascade of care, and the influence of the study on national and international treatment guidelines. However, the ChatGPT version focuses more on potential impacts without capturing all nuances.

Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align significantly with the research team's REF submission, particularly regarding the treatment and epidemiology of Hepatitis C. Both versions discuss the national and international impact on patient outcomes, advancements in the treatment of HCV, and the role of DAAs. There is a high level of accuracy in capturing the essence of the original studies, including benchmarks toward eliminating HCV in line with WHO targets. However, some nuanced details and specific data points are less emphasized in the AI-generated version, hence not receiving the highest rating.

Rater 3 Rating = 4
AAR Justification: The AI-generated impacts closely reflect the outcomes and impacts described in the REF submission. Both versions highlight the national and international influence on HCV treatment guidelines and patient outcomes through the EAP and STOP-HCV initiatives. However, the AI's version is more generalized, lacking some specifics from the REF.

Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated content closely aligns with the REF submission, capturing essential impacts such as enhanced patient outcomes, treatment pathways, and broader public health benefits. The AI content effectively discusses significant areas like national policy influences, data collection contributions, and advancements in clinical practice. However, it lacks the exact reference to the HepCATT study and specific quantitative impacts seen in the REF, such as reductions in liver-related transplants and mortality, which are central to the original submission's claims.

"
"36. Study Title : Improving Homecare Quality in the UK Through Optimized Workforce Planning

Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts closely align with the REF-submitted version, particularly in addressing the critical aspects of workforce optimization, efficiency improvements, and societal benefits such as improved care quality. Both versions highlight the significance of collaboration with Webroster and the development of OptifAI.

Rater 2 Rating = 4 (Very Good)
AAR Justification: ChatGPT aligns closely with REF submission but lacks specific figures and detailed success stories.

Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated version aligned well with the primary themes and objectives presented in the REF 2021 submission, notably reflecting the importance of efficient scheduling in enhancing care quality and organizational outcomes. The emphasis on technological

contributions and workforce impact aligns closely with the original report's focus.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with key themes in the REF-submitted case study, especially in workforce scheduling and operational efficiency improvements. It accurately reflects the collaboration between the University of Nottingham and Webroster Ltd, as well as the development and launch of OptifAI. However, minor discrepancies in specific policy references and public health benefits led to a slight deduction.

"
"37. Study Title : Transforming vaccine policy for pneumococcal disease leading to significant cost savings in the NHS
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated version closely aligns with the research outcomes stated in the REF 2021 submission. Both versions emphasize the significant impact on public health policy, particularly the influence on pneumococcal vaccination policy in the UK and the US. ChatGPT-generated content accurately reflects the contribution of the research towards cost savings for the NHS and the decision-making process within the Joint Committee for Vaccination and Immunisation (JCVI).
Rater 2 Rating = 4: Very Good
AAR Justification: The AI version aligns well with the REF submission in capturing key impacts on UK policy and cost savings but lacks some details, such as the exact economic savings and specific UK and US policy decision-making processes.
Rater 3 Rating = 4: Very Good
AAR Justification: Both versions reflect accurate impacts on UK and US public health policy, specifically regarding pneumococcal vaccination recommendations for older adults. However, the REF submission included more granular references, particularly JCVI policy details and NHS-specific resource savings (e.g., GBP233 million).
Rater 4 Rating = 4
AAR Justification: The AI-generated version closely aligns with the REF submission in capturing significant outcomes such as the UK vaccine policy shift and cost savings for the NHS. It successfully mentions the key influence on policy and public health, although there are minor omissions in nuanced context-specific details, such as interactions with JCVI that were more explicitly detailed in the REF submission.

"
"38. Study Title : UoP32Househistories: A House Through Time: Shaping a flagship TV series to achieve critical and financial success and inspiring the public to engage with house history
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align well with the REF submission's outcomes, particularly in documenting the cultural, social, and economic impacts of the research. Both versions emphasize the contribution to the success of the TV series, ""A House Through Time,"" the public engagement it generated, and its influence on heritage conservation and interest in house histories. However, certain nuanced contributions, such as specific academic roles, stakeholder testimonials, and detailed viewer metrics in the REF version, were underexplored in the ChatGPT version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impact generally aligns well with the core elements of the REF submission, particularly in its focus on public engagement, the success of the TV series, and the role of academic research in shaping the content. Specific impacts, such as the historical research used in A House Through Time (AHTT), are accurately reflected in both versions.
Rater 3 Rating = 4 (Very Good)
AAR Justification: Both versions accurately describe the impact on public engagement, the influence on TV series production, and the academic contributions. The REF submission provides explicit examples, such as collaborations with the BBC and audience reception data, whereas ChatGPT includes similar impacts but without specific examples.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts reflect a strong understanding of the series' contributions to BBC Two's brand and viewership, closely aligning with the original submission's emphasis on critical acclaim and public engagement. The AI output accurately details public engagement activities and acknowledges Ryan's role in enhancing historical context and visual appeal, contributing to the program's recommission.

"
"39. Study Title : Optimising baggage operations at London Heathrow Airport to achieve cost savings for the aviation industry

Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact aligns well with the overall outcomes and impacts described in the REF submission, particularly regarding the focus on optimizing baggage handling at Heathrow and generating significant cost savings for both the airport and airlines. However, there is a slight difference in specificity: while the REF submission emphasizes tangible, real-world outcomes (e.g., reduced short-landed baggage rates, direct financial savings), the ChatGPT version highlights broader, more theoretical impacts (e.g., scalability to healthcare logistics, potential policy implications). Both are correct but differ in scope and detail.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely aligned with the primary research outcomes and impacts described in the original REF submission. For instance, both versions highlighted the optimization of baggage operations, cost-saving initiatives, and service improvements in Heathrow Airport. However, the AI content sometimes simplified technical aspects such as the specific algorithms used for baggage allocation, which were well-detailed in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated report closely aligns with the core aspects of the REF submission, including the focus on baggage operations efficiency, dynamic scheduling, and multi-objective optimization. It captures essential elements such as real-time adjustments and stakeholder concerns, showcasing a strong understanding of the project's primary outcomes.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts closely align with the Heathrow study by highlighting the significance of efficient baggage management, resource allocation, and the economic benefits achieved through optimisation. Both versions discuss the importance of cost savings and passenger experience enhancement through dynamic scheduling, accurately reflecting core project outcomes.

"
"40. Study Title : POWeR - Cost-effective online support for weight management
Rater 1 Rating = 4
AAR Justification: Both the ChatGPT-generated impacts and REF submission are aligned with the core outcomes of the POWeR intervention. Both documents emphasize the project's health impact, scalability, and policy influence. However, the ChatGPT version sometimes provides broader generalizations without the same level of specific detail as the REF submission.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated content aligns well with the key outcomes and impacts of the original REF submission. Both versions highlight the clinical benefits of the POWeR intervention, particularly its success in primary care settings and its cost-effectiveness. They also agree on the importance of the dissemination of POWeR through collaborations with public health bodies and councils. The AI version provides a clear reflection of the project's core impacts, such as sustained weight loss and integration into NHS services. However, some nuanced aspects, such as the iterative process of customizing POWeR for local authorities and the specific details of the contracts signed with Hampshire County Council, Solent NHS Trust, and Redcar and Cleveland Borough Council, were either generalized or omitted in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the outcomes in the REF submission, covering primary objectives like weight management through digital interventions in primary care. Both versions discuss dissemination strategies and target beneficiaries, particularly the NHS and public health stakeholders. However, the AI-generated version lacked some specifics about licensing contracts and policy interactions essential to the REF's practical impact.
Rater 4 Rating = 4
AAR Justification: Both versions align in describing POWeR's core objective to support obesity management through a scalable digital intervention. Both specify the primary care setting and emphasize resource efficiency. However, the ChatGPT version lacks detailed references to specific partners (e.g., Hampshire County Council) and some dissemination pathways noted in the REF submission, affecting full alignment.

"
"41. Study Title : The global impact of Sunderland's football research on scouting, training, and player preparation in elite football
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts closely align with the main themes and findings of the REF submission, especially in areas like the evolution of training methodologies, high-intensity running, and the overall physical and technical demands on elite football players. The AI version captures the core findings accurately but lacks some specific stakeholder engagements and citations that the REF version provides.

Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated research impact aligns closely with the key elements of the REF submission. Both versions accurately emphasize the significant impact of the Sunderland research on football clubs worldwide, particularly in training and player preparation. Specific mentions of high-intensity running, tactical demands, and the research's influence on major clubs like Manchester United, Arsenal, and Barcelona were included in both versions.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated impact analysis aligns well with the core outcomes highlighted in the REF 2021 submission. Both versions accurately reference the significance of Sunderland's research on player physical and technical development, specifically regarding high-intensity demands, tactical insights, and injury prevention in elite football. However, the AI version provides general insights on sports performance without detailing specific sports organizations, such as UEFA and FIFA, where the REF submission offers exact partnerships and direct citations.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT version aligns well with key impacts outlined in the REF submission, especially in emphasizing performance management benchmarks, enhanced training protocols, and global influence across clubs and sports organizations. However, minor discrepancies in specifics such as organization names and exact player statistics affect the direct comparability slightly.

"
"42. Study Title : Adding value to convenience retailing through improved pricing, a new store concept, and capacity building
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the key aspects of the original REF submission. It correctly identifies significant impacts such as changes in store concepts and consumer behavior in convenience retailing. However, some specific details like the names of involved stakeholders and precise numerical outcomes were generalized or omitted.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated version accurately captures many of the key impacts mentioned in the REF submission, including the shift in target demographics, the redefined store concepts, and the effect on sales due to changes in store layout and price management. However, some detailed aspects, such as specific statistical data on sales improvements and award recognitions (e.g., ""Shop of the Year"" awards), were either mentioned briefly or omitted.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the REF-submitted impact in terms of objectives such as enhancing convenience retail operations, improving customer experience, and developing new store concepts. Both versions address similar economic, social, and technological outcomes, with the AI-generated version accurately reflecting key areas like consumer-focused retail adjustments and capacity-building initiatives for store managers. However, while the AI version captures general impact areas, certain nuanced details present in the REF submission are missing.
Rater 4 Rating = 4
AAR Justification: The ChatGPT-generated impacts align with the core research outcomes, notably in consumer behavior's influence on store design, pricing adjustments, and capacity-building benefits. The AI summary accurately represents shifts in retail strategies and customer focus but lacks some finer aspects, like the explicit influence on industry recognition awards and specific sales figures.

"
"43. Study Title : Transforming Evidence-Based Practice in Public Health Through Co-Production and Evaluation
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated content aligned well with the original REF submission in terms of identifying the key impacts on public policy, mental health interventions, and community engagement, as well as the academic contribution to public health. However, minor details such as the integration with specific services (e.g., CREE men's shed project) were not fully captured.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the REF-submitted research impacts, capturing key outcomes such as co-production methodologies and public health interventions. Both emphasize mental health interventions (e.g., the CREE project), suicide prevention, and teenage pregnancy support. However, the REF version contains more nuanced details regarding specific policy changes and their immediate effects on local public health policies.

Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact aligns well with the original REF-submitted content by correctly identifying core impacts on public policy, mental health, and the role of coproduction. It captures significant areas such as mental health support programs, public health policies, and educational interventions through co-production.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts align well with the REF-submitted impacts across most project dimensions, capturing core public health interventions such as mental health, teenage parent support, and suicide surveillance. The AI document accurately reflects the ongoing DCC collaborations in mental health support initiatives (e.g., CREE men's shed), teenage pregnancy programs, and suicide prevention strategies. However, specific project outcomes in the REF, such as DCC's broader use of co-production evaluation to shape practice, are only partially covered by the AI content.

"
"44. Study Title : Well-being and Public Policy
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content largely aligns with the REF 2021 submission, particularly in addressing public policy influence, including the collaboration with ONS and HM Treasury. The well-being focus is maintained across both documents, ensuring key policy discussions are present. However, the AI version occasionally misses nuanced details about specific contributions to UK public health policy and lacks the depth of policy consultation impacts described in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impact for ""Well-being and Public Policy"" aligns well with the main outcomes presented in the REF submission. Both highlight Oswald's contributions to the measurement of well-being and its influence on public policy in the UK, including the integration of well-being metrics into government evaluations, such as HM Treasury's 2018 Green Book. The AI output also covers key collaborations with the ONS and other governmental bodies. However, minor differences in phrasing and depth of specific case study examples slightly reduce the score from Excellent to Very Good.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align well with the REF-submitted version, capturing key elements like the influence on UK policy, public awareness, and productivity at work. Both versions highlight the research's role in informing government policy through the 2018 Green Book and ONS collaboration. However, while the ChatGPT version accurately covers most major impacts, some nuanced contributions to European policy and specific local applications were less detailed.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with core aspects of the REF-submitted case study, particularly in acknowledging the influence on public policy, the UK's focus on well-being measures, and applications in policymaking frameworks. While both versions emphasize the importance of well-being metrics and consultations with policy bodies, minor details like the specific engagements with ONS and Treasury events (2013-2017) were less emphasized in the AI version. Overall, alignment was strong but lacked a few nuances.

"
"45. Study Title : Improving the quality of green infrastructure in towns and cities in the UK
Rater 1 Rating = 4 - Very Good
AAR Justification: Both the AI-generated impacts and the REF submission align well in their focus on green infrastructure, its policy influence, and the environmental and social benefits. Specific examples, such as the creation of Building with Nature standards, were mentioned in both versions. However, the AI version is slightly less detailed regarding the partnerships with local authorities and specific policy changes. The AI-generated content captures the broader impacts, including improvements in planning policies and climate resilience, but misses some precise examples provided in the REF submission, such as the adoption by Newcastle City Council and Homes England.
Rater 2 Rating = 4
AAR Justification: The ChatGPT-generated content aligns well with the REF submission, accurately covering core aspects like green infrastructure improvements and their societal benefits. It mirrors the REF submission by addressing themes such as better planning, biodiversity improvement, and policy influence. However, the REF submission includes additional detail on collaborations (e.g., with Gloucestershire Wildlife Trust) and specific examples of impacts not fully captured in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated version aligns closely with the original case study,

covering most aspects of the stated impact areas, especially around public policy, green infrastructure quality, and stakeholder engagement. However, minor details regarding stakeholder perspectives, especially specific feedback from local authorities, were less emphasized.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align closely with the outcomes and goals outlined in the REF submission, specifically focusing on improvements in environmental quality, community health, and biodiversity within urban development contexts. Both versions address policy impacts and the role of frameworks for guiding infrastructure improvements, though AI-generated content generalizes the socio-economic impacts across broader societal benefits.

"
"46. Study Title : Redressing the state of the stateless: seeking political recognition for Tibet and Kashmir
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated content captures key impacts related to asylum rights for Tibetans and international pressure on India regarding Kashmir. ChatGPT aligns well with the legal and policy impacts presented in the REF submission, such as the asylum cases and U.S. House Resolutions influenced by Kaul's work. The AI-generated version reflects these aspects with strong alignment to core themes. However, some nuanced points like the personal costs experienced by the researchers, particularly related to social media abuse, are less emphasized in ChatGPT.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the key themes and outcomes described in the REF submission. For instance, both versions highlight Anand's contributions to the asylum rights of Tibetans in the UK and Kaul's advocacy for Kashmir's human rights. These reflect an accurate representation of the original submission's legal and political impact.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts closely reflect the main themes of the REF-submitted version, particularly regarding political advocacy and rights for stateless groups. Both emphasize Anand's and Kaul's contributions to asylum rights for Tibetans and international awareness for Kashmiris, aligning well with the REF version's core impacts on policy, public discourse, and advocacy.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align well with the REF submission by accurately identifying key outcomes, such as influencing asylum laws for Tibetan refugees and amplifying international awareness of Kashmir's situation. The rating is ""Very Good"" due to minor variances, such as the emphasis on geopolitical dynamics, which were less pronounced in the AI version.

"
"47. Study Title : Evidence-based enteral feeding practices for very preterm or very low birth weight infants
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated version aligns well with the REF-submitted impacts in major areas, including health, policy, and clinical practice. Key topics such as enteral feeding strategies, the use of human donor milk, and the timing of feeding interventions are consistently mentioned across both versions.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impacts align well with the REF-submitted outcomes in several key areas, particularly in its mention of enteral feeding strategies and their impact on neonatal care. Both versions emphasize the reduction in necrotizing enterocolitis and the use of donor human milk. However, there are subtle differences in how specific evidence-based guidelines are discussed, which brings down the rating from ""Excellent"" to ""Very Good.""
Rater 3 Rating = 4: Very Good
AAR Justification: Both versions align well on major impacts, such as clinical practice improvements, evidence informing feeding guidelines, and reductions in necrotizing enterocolitis (NEC). The ChatGPT version includes thorough references to these primary areas. However, it lacks some specific citations to guidelines like the AAP or CCG seen in the REF submission, which could enhance policy influence.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts align well with the primary themes outlined in the REF 2021 submission, particularly in discussing donor human milk and specific feeding strategies like early trophic feeding and nutrient fortification. Key policy influences, such as those impacting WHO guidelines, are also acknowledged. However, minor discrepancies exist in terminology specificity and the lack of some nuanced policy references.

"

"48. Study Title : Improving Environmental Conservation in East Africa and Beyond
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the majority of outcomes and impacts described in the original REF submission. Both versions mention the core impacts on biodiversity conservation, management strategies, and socio-economic benefits. However, there are some minor differences in the way the AI interprets the importance of policy impact and academic contributions.
Rater 2 Rating = 4
AAR Justification: The AI-generated impacts generally align well with the original REF submission in terms of addressing core outcomes related to conservation management, ex situ conservation, and ecosystem protection. For example, both versions emphasize the successful establishment of the Magombera Nature Reserve and contributions to forest conservation in Tanzania. However, certain nuances like the specific collaboration with local communities and the training programs for stove construction were mentioned in the REF but not emphasized in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the objectives and outcomes stated in the REF submission, capturing key themes such as conservation management and biodiversity. While both versions address conservation drivers and the establishment of Reforest Africa, the AI-generated version missed certain specific details about partnerships and local economic benefits.
Rater 4 Rating = 4: Very Good
AAR Justification: The ChatGPT-generated impact aligns closely with the REF-submitted version, particularly in highlighting key conservation management outcomes, such as the establishment of Reforest Africa and the protection of the Magombera Nature Reserve. The accuracy of location-specific impacts and stakeholder engagement is well preserved in both versions, maintaining consistency with reported project milestones. While most impacts align, there are minor discrepancies in specific data points and geographic details.

"

"49. Study Title : Predicting the properties of materials with first-principles electronic structure software (CASTEP)
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated version captures most of the key elements from the REF submission, including the industrial impact, technological innovation, and academic contributions of CASTEP. However, there are minor inconsistencies in the depth of policy and societal impacts when compared to the detailed real-world use cases of CASTEP in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impact accurately reflects key aspects of the REF submission, particularly in terms of explaining the industrial applications of CASTEP in sectors like semiconductors, pharmaceuticals, and materials engineering. However, there are some minor omissions in the fine details of certain case studies, particularly in terms of how certain algorithms were applied in industry settings.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align well with the core outcomes from the REF submission, reflecting the CASTEP project's contribution to fields like materials science, computational chemistry, and academic research advancements. Both versions mention CASTEP's impact on industries (semiconductors, pharmaceuticals) and its foundational role in computational research through DFT, though the AI-generated impact is more generalized.
Rater 4 Rating = 4 – Very Good
AAR Justification: The AI-generated impact aligns well with the original case study's details, specifically regarding the commercial and academic success of CASTEP in industrial applications. Key aspects, like the use of CASTEP in optimizing material properties and its role in industrial R&D across multinational companies, are reflected with high accuracy. However, some specific industrial use cases and detailed ROI statistics were less prominently highlighted in the AI-generated version.

"

"50. Study Title : Statistical pattern recognition applied to protein crystallisation images in the pharmaceutical industry
Rater 1 Rating = 4 – Very Good
AAR Justification: The ChatGPT-generated impacts closely align with the REF 2021 submission, especially regarding technological advancements in crystallization automation and stakeholder engagement with pharmaceutical companies like AstraZeneca and Merck. Both versions emphasize

automation in image analysis, the significance of MARCO, and the resulting efficiency in drug discovery. However, some specific case examples from industry partners were less detailed in the AI-generated impacts.
Rater 2 Rating = 4
AAR Justification: The AI-generated impacts align well with the outcomes and impacts described in the REF submission, particularly in terms of the technological and economic benefits of automated crystallization. For example, both versions discuss the integration of machine learning and image classification in accelerating drug discovery. However, the REF version contains more precise details regarding specific partnerships (AstraZeneca, Merck), which are only briefly mentioned or generalized in the AI version.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align closely with key aspects outlined in the REF submission, including the importance of automated crystallization and the integration of MARCO software in pharmaceutical workflows. Both versions emphasize the role of AI and automation in drug discovery and address applications across major pharmaceutical companies. The AI-generated version effectively mirrors the critical role of MARCO software as highlighted in the REF submission.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with key aspects of the REF submission, particularly the automation of crystallization processes, involvement of key stakeholders like AstraZeneca, and the increased efficiency for the drug discovery pipeline. Both versions mention advancements in reducing manual labor and improving accuracy in high-throughput experimentation.

"
"51. Study Title : 'Moving beyond one-size-fits-all: Improving Widening Participation through Realist Evaluation methodologies in Northern England'
Rater 1 Rating = 4
AAR Justification: The AI-generated version aligns well with the REF-submitted version. It captures the core aspects of widening participation (WP) and the role of Realist Evaluation (RE) in improving access to higher education. Both versions address the impact on disadvantaged communities and recognize the role of outreach staff in community settings. However, the AI version sometimes lacks the specific depth in examples of RE practices.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the REF-submitted version, reflecting key themes such as the use of Realist Evaluation (RE) in Widening Participation (WP) and the involvement of Higher Education Progression Officers (HEPOs) to support underrepresented communities. Both versions discuss how the program facilitated access to higher education for students from disadvantaged areas. However, the REF submission contains more granular detail, particularly about the geographical focus on West Yorkshire and specifics of how the interventions were shaped.
Rater 3 Rating = 4
AAR Justification: The ChatGPT version largely aligns with the outcomes highlighted in the REF submission, accurately representing key impacts on policy, educational practices, and outreach methodology. However, certain nuanced aspects, such as specific program theory contributions and stakeholder roles, were better detailed in the REF version.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated version aligns closely with key components highlighted in the REF submission, accurately referencing core aspects of widening participation (WP), community outreach, and the Realist Evaluation (RE) approach used to assess WP's efficacy. Both versions consistently mention the emphasis on underrepresented communities and the context-driven outcomes essential to WP's success. However, some nuanced details regarding the specific geographic and institutional partnerships were slightly generalized in the AI version.

"
"52. Study Title : Challenging monolithic conceptualisations of English for learning, teaching and assessment: The Changing Englishes online course
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the REF submission's key outcomes, such as promoting plurilithic views of English and influencing ELT practices globally. Both versions address the importance of linguistic diversity, learner inclusivity, and the reevaluation of language norms in teaching contexts.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impacts aligned well with the major themes of the REF submission, such as the focus on promoting plurilithic views of English and challenging the dominance of monolithic norms in ELT. For instance, both versions accurately reflect the shift in teacher beliefs and practices based on the Changing Englishes course. However, there were

minor discrepancies in specificity, such as missing direct citations or statistics related to user participation in the course.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the key outcomes in the REF-submitted version, particularly in representing the plurilithic approach to English, with strong references to both the theoretical foundation and the educational applications. However, the REF submission includes specific impact cases in certain geographic and cultural contexts, which the AI version does not consistently cover.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated version accurately captures the plurilithic framework and the impact on English Language Teaching (ELT) by addressing the significance of non-standard varieties and the socio-political influences on language norms. However, minor deviations in emphasis on specific qualitative testimonials from educators limit a perfect alignment.

"
"53. Study Title : Robert Paul and the Origins of British Cinema
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT version captures many of the core impacts presented in the REF submission, including technological advancements, cultural contributions, and commercialization potential. For instance, both versions align on impacts related to early cinema technologies and their influence on advertising and cultural preservation.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated version correctly identifies the technological, cultural, and historical contributions of Robert Paul, particularly his innovations in cinema technology and the preservation of early films, which are aligned with the REF-submitted case study. However, some detailed aspects, such as specific exhibitions or particular film discoveries, are not fully mirrored in the ChatGPT version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated content aligns closely with the outcomes and impacts described in the REF 2021 submission. The AI version captures the technological, cultural, and societal contributions of Robert Paul's work, including his innovations in film, advertising, and education, and accurately reflects his role in early British cinema.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts effectively capture many core aspects of the REF-submitted version, such as the discovery and restoration of Robert Paul's films and the project's influence on the recognition of early British cinema. This includes broad alignment on the historical and technological relevance of Paul's work and his contributions to British film culture. However, the AI version lacks some specific institutional references and details on individual films identified and restored.

"
"54. Study Title : The Care and Management of Gout in Primary Care
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content accurately reflects the core themes of the REF submission, including the impact on gout management guidelines (EULAR and BSR). The ChatGPT version addresses key aspects like rising gout prevalence and the need for improved patient management through updated clinical guidelines. However, specific examples, such as the role of NICE and CKD management, were better detailed in the REF version.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated content aligns well with the major outcomes and impacts discussed in the REF submission, such as the impact on gout management guidelines and public health interventions. Both versions emphasize the increasing prevalence of gout, the need for urate-lowering therapies, and the significance of gout comorbidities like cardiovascular disease and CKD. The AI output accurately reflects key aspects like regional health disparities and treatment protocols, but sometimes lacks the precise, granular detail found in the REF submission, particularly regarding specific guideline mentions like those from NICE.
Rater 3 Rating = 4
AAR Justification: The ChatGPT-generated content effectively aligned with most of the outcomes in the REF submission, capturing critical impacts on national guidelines, awareness efforts, and the importance of early urate-lowering therapy. However, the AI version missed certain specifics related to how the findings were applied directly within NHS practices, which are crucial for comprehensive alignment.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated content closely aligns with the core outcomes outlined in the REF submission, including improvements in gout management guidelines and impacts on patient care protocols. Specific references, such as the inclusion of cardiovascular risk screening and

chronic kidney disease (CKD) management, were present in both versions.

"
"55. Study Title : From Victims to Actors: Shifting the Policy Paradigm to Value Children's
Contribution in Disaster Risk Management
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts largely aligned with the outcomes of the original
REF submission, especially in highlighting children's involvement in disaster risk management,
their agency, and the influence on policies. Both versions covered the core themes of children's
empowerment and shifts in disaster management strategies.
Rater 2 Rating = 4 - Very Good
AAR Justification: Both the AI-generated and REF submissions show strong alignment in key
aspects of the research impacts, particularly around children's agency in disaster risk
management, influencing policy changes, and reshaping education in flood resilience. However,
there were some nuanced differences. For example, while both versions emphasized the importance
of children in disaster recovery, the AI-generated version did not capture all of the specific
case study examples (e.g., the inclusion of children in the Surrey County Council's flood
response efforts).
Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated content accurately reflects the primary themes and outcomes
described in the REF submission, such as the shift in policy and practice towards recognizing
children's active roles in disaster risk management. However, certain nuanced aspects, such as
specific references to UK policies and direct quotes from child participants, are less
explicitly highlighted in the AI-generated content.
Rater 4 Rating = 5
AAR Justification: Both versions accurately align with the core objectives and outcomes. The REF
submission places a strong emphasis on children as proactive contributors to disaster
management, impacting policies across multiple regions. The AI-generated impact similarly
underscores children's empowerment and policy influence but simplifies specific stakeholder
engagement and geographic scope details.

"
"56. Study Title : Improved Crab Fisheries Management Benefits Coastal Livelihoods in Brazil
Rater 1 Rating = 4 - Very Good
AAR Justification: Both the ChatGPT and REF versions align well on the core outcomes, such as
policy changes in Brazil's fisheries management, especially the prediction tools for crab mass
mating and the socioeconomic impacts on artisanal fishers. Both also discuss legislative impacts
and the broader economic implications. The AI-generated version largely follows the direction
and thematic focus of the REF-submitted report but lacks some finer details provided in the REF
submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact assessment aligns well with the original case study
in capturing the key elements of policy change, sustainability of the crab fishery, and
socio-economic benefits. Both versions emphasize the role of accurate mass-mating predictions in
transforming fisheries management in Brazil. However, there are subtle differences in phrasing
and focus, leading to slight misalignments.
Rater 3 Rating = 4 - Very Good
AAR Justification: Both versions capture the core impact on fisheries management and the
socio-economic benefits to coastal communities. The AI-generated version aligns well with the
REF submission in highlighting policy influence, socio-economic impacts on crab fishers, and
environmental sustainability. However, the REF version provides more specific details on
governmental engagement and legislative actions.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts align substantially with the original REF
submission. Both versions highlight the significance of environmental and socio-economic
benefits stemming from improved fisheries management practices, especially in reducing income
loss and enhancing sustainability in crab harvesting. Notably, both versions underscore the
policy shift towards targeted fishing bans based on accurate predictions. However, the REF
version contains finer details on stakeholder interactions, specifically government agencies'
roles in facilitating policy adoption.

"
"57. Study Title : Influencing Organisational Strategy to Support Responsible Business Practice
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts largely align with the REF-submitted version. Both
versions reflect the significant contributions to responsible business practices, public policy

influence (e.g., Scottish Business Pledge, CAN B status for Edinburgh), and cross-sector collaborations. Key impacts such as policy shifts and community engagements are consistent in both versions.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the original research impacts, especially in core areas such as responsible business practices and stakeholder collaboration. Both versions discuss the impact on public policy (Scottish Business Pledge) and the promotion of business responsibility through forums like the Responsible Business Forum. However, certain nuances, such as the specific involvement of Edinburgh Napier University as a knowledge partner in Scotland CAN B and the impact of specific events like the Charity Board Initiative, were covered more comprehensively in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the outcomes described in the REF submission, particularly in areas such as policy influence, societal impact through community engagement, and economic benefits from partnerships.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated content closely aligns with the REF 2021 submission's impact areas, such as the role in fostering business-community engagement, policy influence, and supporting environmental and social goals through responsible business practices. Specific initiatives like the Scottish Business Pledge and CAN B designation were accurately captured.

"
"58. Study Title : Billmonitor: predicting the best mobile phone contract for businesses and individual users
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated content closely mirrors the economic impact described in the REF submission, particularly in identifying cost savings for consumers and businesses through the Billmonitor service. However, the AI version lacks the same specificity, such as the exact figures for the total savings (e.g., GBP37 million identified savings for NHS and other public sector organizations), but it accurately captures the broader financial benefits. The AI-generated version also correctly identifies the economic and academic domains of impact, but it omits some finer details related to stakeholder engagement and press endorsements.
Rater 2 Rating = 4 - Very Good
AAR Justification: Both the AI-generated and REF-submitted versions are well-aligned in terms of the core aspects of the research, particularly the predictive algorithm for mobile usage and its economic implications. The AI report captures the central mechanism of the research but expands it with broader applications in cross-disciplinary contexts (e.g., statistical modeling in finance and genetics). The AI-generated report also discusses the theoretical aspects, providing a sound alignment with the original REF submission. However, it lacks some specific details such as the partnerships with Ofcom or specific testimonials from users, which are more pronounced in the REF version.
Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts capture key economic aspects and emphasize the significant cost savings for both private users and SMEs, as highlighted in the REF submission. Both versions align in detailing the primary economic benefits and the algorithmic innovation of Billmonitor in predicting user needs. However, the REF version includes additional context regarding the involvement of public sector organizations (NHS), which ChatGPT partially captured but not in as much detail.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts largely align with the detailed outcomes from the REF submission, particularly on financial savings and the value delivered to businesses and public entities. It correctly captures the impact on economic aspects and the utility for SME and individual users, which are central themes in the REF submission. However, some nuances regarding policy influence and indirect market impacts are less explicit in the AI-generated version.

"
"59. Study Title : Mediating Modern German: reaching new and diverse audiences through translation, engagement, and performance
Rater 1 Rating = 4 - Very Good
AAR Justification: Both versions align well in describing the impact on public engagement, translation, and cultural outreach. The ChatGPT version accurately reflects core aspects such as the expansion of German poetry to new audiences and collaboration with festivals and organizations, as mentioned in the REF submission. However, some nuances, like the specific collaboration with the British Library and the details about Leeder's role in poetry competitions, are less explicit in the ChatGPT version.

Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated research impact aligns closely with the REF submission in several key areas, particularly in terms of public engagement, translation efforts, and the role of German poets in cultural exchange. The AI version mentions the broad goals of reaching new audiences and influencing cultural practices, which are central themes in the REF submission. However, while the AI-generated impacts address translation and audience engagement, they lack specific details on individual achievements (such as awards or judging competitions) that are highlighted in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns closely with key impacts outlined in the REF 2021 submission. It accurately highlights public awareness and engagement, increased access to German poetry, and Leeder's work with multiple organizations and festivals. However, certain nuanced details regarding specific collaborations and accolades, such as awards and participation in distinct events, were less emphasized in the AI version, which affects the overall alignment.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts align closely with the outcomes mentioned in the REF submission, such as public engagement in German poetry and collaboration with arts organizations. However, specific awards and events related to Leeder's role, as mentioned in the REF, are less detailed in the AI-generated version. The AI successfully captures the broad impact on public awareness and collaboration outcomes.

"
"60. Study Title : OMass Therapeutics: New technology for drug discovery with economic benefit to the UK
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated version closely matches the REF submission in many core areas. Both versions focus on key impacts such as technological advancements in mass spectrometry, economic contributions through the creation of OMass Technologies, and contributions to healthcare through drug discovery. The AI version captures these aspects well, though it is not as detailed in certain sections like specific company milestones.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated impacts largely align with the REF-submitted version in discussing OMass' drug discovery spinout, highlighting technological innovations (e.g., mass spectrometry) and economic impacts. However, some nuanced details on specific drug pipeline advancements and UK-specific financial investments were more deeply outlined in the REF submission.
Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated content largely aligns with the REF 2021 submission, capturing key elements like the establishment of OMass Technologies, technology's unique contribution to drug discovery, and economic impacts. Both versions mention patent portfolios, the rapid identification capabilities of the platform, and the creation of job opportunities.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated content closely aligns with the main impacts described in the REF submission. It reflects the creation of OMass, the development of novel mass spectrometry technology, and significant milestones such as patent generation, company spin-off, and pharmaceutical partnerships. However, some minor contextual details around specific investor relationships and awards were less emphasized, which slightly impacts the alignment rating.

"
"61. Study Title : Living With Feeling: Transforming Understandings of Emotional Health
Rater 1 Rating = 4
AAR Justification: The AI-generated content closely aligns with the detailed outcomes described in the REF submission. Both versions emphasize the contributions to emotional health research, public engagement through media (e.g., BBC), and collaboration with institutions like the Royal College of Nursing (RCN). The AI version accurately reflects these impacts but lacks a few intricate details about the public's reception and specific collaborations mentioned in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the core outcomes described in the REF submission. The key themes, such as public understanding of emotions, engagement with media (e.g., BBC), and the historical insights into emotions like altruism, anger, and the 'stiff upper lip,' are accurately captured. However, the AI report lacks some specific details related to partnerships and precise outcomes of public exhibitions.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts closely align with the main objectives of the

original research, especially in detailing emotional health's historical and societal influences. Key impacts, such as public engagement through media and educational programs, are well-represented, though some nuanced details on engagement with specific stakeholders are less emphasized.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts closely align with the original outcomes described in the REF submission, particularly regarding the project's influence on public understanding and mental health perspectives. The AI-generated content successfully captures the essence of the project's impacts on public engagement through media and partnerships with organizations like BBC and RCN. However, it occasionally generalizes specific programs and stakeholder interactions, omitting some key details present in the REF submission, such as individual series or event outcomes.

"
"62. Study Title : Litigation as a Tool to Support Social Change: Indigenous Peoples, Human Rights, and Legal Empowerment
Rater 1 Rating = 4: Very Good
AAR Justification: The AI-generated research impacts align closely with the outcomes described in the REF 2021 submission. Both emphasize the influence of litigation on indigenous rights, policy changes, and international norms. For example, both versions address the impact on the Namibian case and international NGOs such as OSJI. However, the AI version captures slightly broader domains (e.g., educational domain), while the REF focuses more on the specific cases.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the core elements of the research described in the REF submission, such as litigation for indigenous rights and its societal impact. Key components like the Namibia and Kenya cases and the role of international NGOs are accurately mirrored. However, it misses some intricate procedural details and stakeholder roles that are present in the REF submission.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated version aligns closely with the REF submission by accurately identifying core impact themes, such as indigenous empowerment through litigation, global policy influence, and the role of advocacy. Key points, including the focus on indigenous community empowerment, litigation as a tool for advocacy, and policy impacts on institutions like the UN, are well-matched. Some specific litigation case details, such as the Batwa in Congo, are missing in the AI version, resulting in a slightly reduced rating.
Rater 4 Rating = 4: Very Good
AAR Justification: Both versions accurately address the primary impact areas, including legal empowerment, indigenous rights, and influence on UN policies. The REF version offers deeper specifics in terms of cases and community impacts, while the AI-generated report reflects similar themes with broader examples. However, slight variations exist in how these impacts are contextualized for long-term policy influence.

"
"63. Study Title : Accelerating the development of medicines for children through an open-access excipient database
Rater 1 Rating = 4 - Very Good
AAR Justification: Both the REF-submitted and ChatGPT versions align in terms of the central theme of developing an excipient database and its impact on pediatric medicine. ChatGPT accurately captures the key impacts, including regulatory influence and clinical applications (e.g., support of pediatric medicine formulation and EMA collaboration). However, minor details, such as specific examples of excipient inclusion (e.g., Proveca's example in REF) are less emphasized in the ChatGPT version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated version accurately captures the core impact described in the REF submission, particularly the creation of the STEP database and its use by pharmaceutical companies and regulatory agencies. The ChatGPT version misses some specific details, such as precise statistics on global user engagement and company-specific endorsements, but generally aligns well with the main themes and outcomes.
Rater 3 Rating = 4 (Very Good)
AAR Justification: Both versions strongly align on the main goal of promoting safe pediatric formulations through the STEP database. The REF submission offers detailed examples (e.g., usage by major pharmaceutical companies) and specific regulatory endorsements that substantiate the project's impact. ChatGPT's version captures the broader objectives, like enhancing pediatric medication safety and aligning with regulatory needs globally, yet occasionally lacks specific mentions of certain pharmaceutical examples.
Rater 4 Rating = 4 (Very Good)

AAR Justification: The ChatGPT version aligns closely with the original REF submission, accurately identifying the STEP database's objectives and the broad impacts on pharmaceutical companies, regulatory bodies, and healthcare professionals. While specific references to companies like Eli Lilly and Proveca were absent, the general alignment in highlighting the database's global accessibility and role in enhancing pediatric drug development was maintained.

"

"64. Study Title : Pluralistic Evidence for Successful Policymaking about Reactive Systems

Rater 1 Rating = 4 (Very Good)

AAR Justification: The ChatGPT-generated research impacts align well with many of the outcomes described in the original REF 2021 submission. Both versions discuss the importance of information quality (IQ) and its cross-sectoral impacts, particularly in healthcare and data governance. They address how pluralistic evidence informs policymaking and how the research shapes international methodologies in health evidence assessment.

Rater 2 Rating = 4 - Very Good

AAR Justification: The AI-generated version aligns well with the REF submission in several core areas, particularly with respect to the contribution of evidence plurality in public policy and healthcare methodologies. Both versions acknowledge the importance of evidence mechanisms in policymaking. However, while ChatGPT captures most of the significant contributions (e.g., changes to NICE and IARC guidelines), it does not reflect the same depth regarding the specific details of the consultation processes with UK agencies and international bodies (e.g., involvement with NHS and NICE).

Rater 3 Rating = 4 (Very Good)

AAR Justification: The ChatGPT-generated version aligns closely with the REF submission in core impacts on policy, ethical frameworks for AI and data, and healthcare evidence assessment. Both versions address changes in evidence frameworks in NICE and IARC; however, minor discrepancies exist in specificity.

Rater 4 Rating = 4: Very Good

AAR Justification: The AI-generated impacts align well with the main themes of the REF 2021 case study, notably in the areas of public policy influence and ethical frameworks in data science and AI. The AI version accurately mirrors the central themes of pluralistic evidence and the ethical challenges in policy settings. However, certain highly specific procedural changes, such as the influence on NICE and IARC frameworks, are less explicitly articulated.

"

"65. Study Title : Reshaping professional heritage practice and changing understanding of heritage in the UK and internationally

Rater 1 Rating = 4 (Very Good)

AAR Justification: The AI-generated content closely aligns with the REF submission in terms of discussing the impact of heritage on international policy, conservation practices, and stakeholder engagement. It accurately mentions the involvement of UNESCO and other global bodies in rethinking conservation approaches. However, it lacks some specific examples of workshops and co-created events highlighted in the REF submission.

Rater 2 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts accurately reflect many of the outcomes described in the REF submission, such as contributions to international heritage policy (e.g., UNESCO), involvement in workshops, and impacts on public engagement through exhibitions. However, certain nuanced details, especially regarding specific workshops and organizational changes, were less precise. The alignment is high but not perfect.

Rater 3 Rating = 4 - Very Good

AAR Justification: The ChatGPT-generated impacts align well with the core aspects of the REF-submitted case study, particularly in capturing themes such as sustainability, the integration of cultural and natural heritage, and the engagement with international heritage policy organizations. The AI-generated version highlights significant aspects, like cross-disciplinary impacts, though with slightly less emphasis on specific partnerships and workshop activities compared to the REF submission.

Rater 4 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts align well with the primary themes and outcomes noted in the REF submission, including the promotion of integrated heritage management practices and the importance of future-oriented conservation. The ChatGPT version references similar conservation goals and acknowledges the project's impact on policy and practice, as seen in discussions on natural and cultural heritage preservation methods. However, certain nuanced aspects, like the in-depth role of international advisory boards and workshops' direct impact on specific organizations, are not covered in full detail.

"

"66. Study Title : Shaping the legal framework for Brexit
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align closely with the outcomes described in the REF submission. Both versions discuss the influence of the legal arguments on landmark cases like Miller I and Wightman, as well as the legal frameworks surrounding Article 50 and Brexit decision-making. The core legal impact is consistently represented in both.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impact aligns closely with the REF submission, accurately reflecting key aspects like the influence on public policy (e.g., the impact on Article 50 debates). Both versions discuss the legal framework around triggering Article 50 and the landmark Miller case. However, the AI-generated version lacks specific details on some stakeholders' reactions and contextual elements like the timing of the legal challenges.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the major themes in the REF 2021 submission, such as impacts on parliamentary engagement and legal frameworks. Both versions emphasize the role of academic insights in influencing legal proceedings, especially around Article 50. However, minor variances exist in terminology and nuances, impacting complete alignment.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the core legal implications identified in the REF submission, particularly in their coverage of landmark judicial cases such as Miller I and Wightman. These cases are correctly highlighted in the AI content as pivotal to establishing parliamentary authority in Brexit proceedings. However, certain nuanced legal details and the broader political discourse surrounding these rulings, present in the REF, were less emphasized in the AI version.

"
"67. Study Title : Bristol's materials research is keeping the UK's fleet of nuclear power stations safe and operating.
Rater 1 Rating = 4
AAR Justification: The AI-generated content generally aligns well with the original REF 2021 submission, particularly in its discussion of material degradation in AGR nuclear reactors, operational life extension, and impact on energy infrastructure. However, while both the AI version and the REF version focus on crucial aspects like creep, oxidation, and material degradation, the AI content sometimes lacks specific examples or in-depth case studies mentioned in the original REF submission (e.g., the detailed safety case for Hunterston B Reactor).
Rater 2 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impact largely aligns with the core themes from the REF submission. It accurately reflects the project's focus on extending the lifespan of Advanced Gas-cooled Reactors (AGRs), emphasizing contributions to UK electricity generation, carbon emissions reduction, and safety case development. The AI-generated report correctly identified key technical aspects like the analysis of Type-316H steel and core graphite degradation, which were pivotal in the original submission.
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated content aligns well with the research team's submission, capturing the main contributions to energy security, economic benefits, and safety in nuclear power operations. Both versions recognize the importance of UoB's research in extending the lifespan of AGR reactors and its significance for EDF and the UK's energy grid stability. However, the REF submission includes specific safety case examples, such as the role in extending Hunterston B Reactor 3's operating life, which is not explicitly detailed in the AI-generated content.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT version aligns well with the REF submission, covering core aspects such as nuclear safety, material degradation, and impact on UK energy security. Both versions emphasize the life extension of AGRs and the role of material analysis.

"
"68. Study Title : Radical advance in treating age-related macular degeneration leading to global impact in prevention of blindness
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated research impacts closely align with the outcomes and impacts described in the REF submission, particularly in terms of health and economic impacts. Both versions mention the cost-effectiveness and efficacy of bevacizumab and ranibizumab in treating wet AMD, and the large-scale savings for healthcare systems. The description of the research's impact on clinical guidelines and treatment practices is accurate and consistent in both cases.

Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts closely align with key elements of the original REF 2021 submission. The ChatGPT version accurately reflected major impacts on healthcare policy, particularly regarding the influence on NICE guidelines and global changes in clinical practices due to the IVAN trial, similar to the REF case study. However, there were some nuanced differences in phrasing or emphasis, such as ChatGPT focusing more broadly on economic impacts globally than in specific countries mentioned in the REF submission.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated report closely aligns with the REF 2021 submission, accurately reflecting major outcomes and impacts, particularly on health policy and cost-effectiveness analysis. The AI version adequately captures the study's influence on clinical guidelines and international healthcare practices, though it occasionally lacks detail on specific stakeholder engagement as covered in the REF report.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the key outcomes detailed in the REF submission. Specific aspects such as the clinical guidance changes and cost-effectiveness advocacy based on the IVAN trial are reflected accurately in the AI summary. However, some nuances related to stakeholder engagement and specific national responses are less developed.

"
"69. Study Title : Rolling programme of research, centred on the National Joint Registry, to improve the outcomes of hip and knee replacements worldwide
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the REF submission, particularly in recognizing key impacts such as the discontinuation of metal-on-metal implants, influence on healthcare policy, and improvements in clinical practices. The AI-generated content accurately reflects the study's effect on surgical choices, patient outcomes, and regulatory recommendations across the UK and internationally. However, it lacks some country-specific statistics and precise regulatory details that are present in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the REF-submitted impacts in terms of core outcomes, particularly in areas such as the identification of failing implants, changes in policy, and clinical practices concerning metal-on-metal hip replacements. Both versions highlight the impact on public health policies, particularly the reduced use of unsafe implants across multiple countries. However, the AI content occasionally lacks the specificity seen in the REF submission regarding regulatory actions and detailed figures on revision procedures.
Rater 3 Rating = 4
AAR Justification: Both the AI-generated and REF-submitted impacts largely align in their description of the study's influence on clinical guidelines and policy changes, especially regarding the use of safer prosthetic materials. However, some contextual depth around regulatory decisions cited by the REF team, such as FDA and NICE guidelines, was not as deeply matched in the AI-generated version.
Rater 4 Rating = 4
AAR Justification: The ChatGPT-generated impacts align closely with the REF submission, addressing key areas such as public health, cost-effectiveness, and device safety. However, some regulatory and organizational details from the REF submission, like specific guidelines issued by NICE and MHRA, are not explicitly covered in the AI-generated impacts.

"
"70. Study Title : Reducing breast and ovarian cancer occurrences in women at high risk
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align closely with the primary impacts outlined in the REF submission, such as enhancing public health through improved breast cancer risk prediction and influencing healthcare guidelines. However, while both versions reflect accurate impacts on healthcare practices and risk stratification, the AI-generated report occasionally generalized on policy impacts without citing specific guidelines or policy adoption cases as in the REF submission.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the core objectives and achievements of the REF-submitted version. Both versions emphasize the BOADICEA tool's effectiveness in cancer risk prediction and its integration into clinical guidelines for breast and ovarian cancer screening. The AI-generated content, however, lacked some specifics found in the REF version, such as the endorsement by certain international organizations (e.g., NICE and NHS). Nevertheless, the alignment remains strong, particularly in illustrating the tool's significance in patient decision-making and personalized healthcare.

Rater 3 Rating = 4
AAR Justification: The AI-generated impacts closely align with the original REF 2021
submission's impacts, accurately reflecting BOADICEA's clinical, policy, and global influences.
The ChatGPT version includes broad mentions of healthcare applications, empowerment in
preventive decisions, and global guideline integration. However, minor omissions, such as
specific statistics on mortality reduction, slightly reduce complete alignment.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align closely with the REF submission,
accurately reflecting the project's core aims of personalized risk prediction for breast and
ovarian cancer. It captures the tool's utility in clinical practice, referencing its use in
personalized medicine and integration into genetic counseling. Notable alignment is observed in
the emphasis on health outcomes, empowerment of women, and support for clinical decision-making.

"
"71. Study Title : The Haydn Scale: Changing policy and practice for improving pupil behaviour
in schools
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns closely with the key areas outlined in the
REF submission, including policy influence, teacher training improvements, and classroom
management frameworks. Notably, both versions cover the adoption of the Haydn Scale in
educational contexts and its international reach. However, minor discrepancies exist, especially
in nuanced policy implications.
Rater 2 Rating = 4 (Very Good)
AAR Justification: Both versions largely align in reflecting the impact on policy and teacher
education, focusing on pupil behavior and classroom climate. AI-generated impacts accurately
emphasize the relevance of the Haydn Scale in teacher training and behavioral policy, mirroring
core aspects highlighted in the REF version, such as teacher testimonials and DfE policy
influence. However, the AI version does not fully capture the long-standing engagement with
Ofsted and the Bennett Review, which were pivotal in the original submission's influence
narrative.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with key aspects of the REF-submitted
content, specifically capturing the central theme of using the Haydn Scale for improving
classroom management. It accurately reflects the influence on teacher training and policy impact
in UK schools. However, there are minor discrepancies in phrasing and depth in the AI version
compared to the original submission.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align well with the REF 2021 submission in
terms of addressing policy influence and educational improvement, particularly around classroom
climate and behaviour management. Both versions discuss Haydn's contributions to the Bennett
Review and policy shifts in UK schools, reflecting accuracy in contextual details. However, the
REF submission provides more specific instances, such as DfE citation frequency, which enhances
policy accuracy.

"
"72. Study Title : ""Definitive demonstration of the adverse cardiovascular effects of air
pollution leads to policy change at the local, national, and international level""
Rater 1 Rating = 4 - Very Good
AAR Justification: Both versions align closely in highlighting the research's impact on air
pollution's cardiovascular effects and its influence on policy. The AI-generated impact aligns
well with the REF submission's focus on air quality and health. However, the AI version
sometimes generalizes the impact on specific domains, which is highly detailed in the REF
submission.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the core outcomes in the REF
submission, particularly in linking air pollution to cardiovascular health risks and influencing
policy discussions. However, the AI version lacks some precise references found in the REF, such
as media coverage, named collaborations, and events influencing policymakers directly. Overall,
it maintains a close thematic match, accurately reflecting the impact on public health, but
without some finer details.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts accurately reflect key areas highlighted in the REF
2021 submission, such as the link between air pollution and cardiovascular health, policy
advocacy, and public awareness. Both versions emphasize impacts on public health policies and
highlight advocacy efforts, though the AI version occasionally generalizes where the REF
submission provides specific examples (e.g., UoE's involvement with BHF campaigns).

Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with the key impacts discussed in the REF submission, such as the association between air pollution and cardiovascular health and the translation of these findings into public policy and awareness campaigns. Both versions emphasize the effects of particulate matter (PM2.5) on health and highlight the contribution to air quality standards globally.

"
"73. Study Title : Navigating Inclusion in International Peace Processes
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated content effectively aligns with many key impacts specified in the REF 2021 submission, particularly in domains such as influencing international peacebuilding policies, supporting inclusivity, and enhancing global frameworks for women's representation in peace processes. However, some specific details, such as references to certain UN Security Council Resolutions and OECD legal frameworks, are missing, affecting precise alignment. The ChatGPT version demonstrates thorough understanding and alignment in a broad sense but could benefit from direct inclusion of specific resolutions and organizational references.
Rater 2 Rating = 4 – Very Good
AAR Justification: The AI-generated impact description aligns well with the core objectives and key outcomes outlined in the REF submission, capturing the role of the PSRP in promoting gender inclusion in international peace processes. Specific legislative influences, such as UNSCR 2242, are included in both versions, showing a clear alignment. However, the ChatGPT version occasionally generalizes certain legislative contributions without detailing specific articles or clauses mentioned in the REF version, which impacts the precision of alignment. Overall, the alignment is high but could improve with more granular policy references.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align well with the outcomes described in the REF submission, capturing the primary goals of promoting inclusive peace processes, contributing to international policy frameworks (UNSCR, OECD), and supporting gender inclusion. Both versions cover the project's impact on UN policies and initiatives on women's roles in peace processes and the international legal and political frameworks influenced by the PA-X Database. However, certain nuanced references to specific UN recommendations (e.g., UNSCR 2242) and the collaborative work with OECD are only partially covered in the ChatGPT version, limiting a full alignment rating.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated version aligns closely with the REF 2021 submission, accurately reflecting key impacts such as policy influences on UNSCR 2242 and UNSCR 1325 and the collaboration with UN Women for gender inclusion. It also addresses the OECD recommendations on integrating peacebuilding into development, showing a broad understanding of specific policy contributions. However, some nuances on targeted policy instruments in different contexts are slightly generalized in the AI content.

"
"74. Study Title : Documenting and Protecting Survivors of Torture and Ill-Treatment Living in Poor Communities
Rater 1 Rating = 4 (Very Good)
AAR Justification: Both the REF 2021 and ChatGPT-generated versions align well on major impacts such as the influence on human rights documentation practices, particularly regarding torture survivors in low-income regions. The REF submission provides concrete examples of organizational changes, such as the Independent Medico-Legal Unit's (IMLU) strategic adjustments and specific tools like the mobile app for documentation. While the ChatGPT version accurately represents the general impacts, it lacks granular detail on individual organizations' responses and specific activities that were instrumental in achieving these impacts.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the REF submission's documented impacts, particularly regarding the project's influence on human rights documentation practices, technology adoption (e.g., mobile app for torture documentation), and policy development across several countries (e.g., Kenya, Tunisia, Morocco). Both versions recognize key organizations like IMLU and highlight the app's role in broadening access to documentation through community involvement. However, the AI version lacks some specific instances mentioned in the REF, which slightly limits full alignment.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated version closely aligns with the REF submission, covering key impacts on human rights documentation, awareness, and policy. It accurately captures the essence of the research's objectives and outcomes, such as influencing human rights

organizations' methods and advocating for torture survivors.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts aligned closely with the primary impacts of the REF-submitted study, such as influencing public policy, increasing awareness, and refining human rights documentation methods, especially for marginalized communities. The AI report correctly referenced similar organizations and regions impacted, though some details were generalized rather than project-specific.

"
"75. Study Title : Transforming genomic selection in commercial breeding programmes for pigs, dairy goats, and poultry
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts largely align with the REF 2021 submission in terms of reflecting the transformative effects on genomic selection for productivity and efficiency gains across livestock. Both versions detail increased accuracy in selection traits, sustainability benefits, and economic outcomes. However, the REF version provides richer quantitative insights specific to species, such as the economic projections for Genus' global operations and targeted genetic traits, which could benefit from further mention in the AI-generated impacts.
Rater 2 Rating = Rating: 4 (Very Good)
AAR Justification: The AI-generated impact analysis aligns well with the REF submission, accurately reflecting the central impacts on productivity and sustainability within the specified species (pigs, dairy goats, poultry). The AI content accurately addresses the increase in breeding program efficiency, genomic selection benefits, and the associated environmental contributions. However, the AI version could benefit from additional specificity on financial and productivity metrics, such as increased piglet productivity and global economic gains. The analysis is closely aligned with the REF's goals and structure, though with fewer details on immediate financial impacts and partner-specific results (e.g., improvements noted by Genus and Yorkshire Dairy Goats).
Rater 3 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align closely with those documented in the REF submission. Both versions reflect advancements in genomic selection and their implications across species (pigs, goats, poultry). AI content accurately mirrors the original's impact on productivity, selection accuracy, and breeding program efficiency but lacks the precision in financial impact for Genus and YDG in productivity gains and profitability improvements, respectively.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the REF 2021 submission in key areas of genetic selection improvements, economic benefits, and global reach. It addresses major species-focused advances in productivity for pigs, goats, and poultry and covers genetic gain metrics similar to REF's claims of increased accuracy and economic output. However, it is somewhat less precise in quantifying individual gains and lacks detailed mention of collaborative roles that the REF version highlights.

"
"76. Study Title : Employing polymer physics to improve gluten-free bread structure
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the key impacts detailed in the REF submission, especially concerning the role of viscoelastic materials in gluten-free bread production and improvements in manufacturing consistency, as well as polymer physics insights applied in food science. The AI mentions multiple applications in food industries and manufacturing, closely matching the emphasis on improved gluten-free product quality and stability seen in the REF.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the specific impacts highlighted in the REF submission, such as improving bread consistency through polymer research and addressing instability in gluten-free products. It acknowledges the scientific basis provided by Prof. Morozov's expertise on polymer instability in practical applications. However, while the AI version accurately mentions consumer relevance and economic outcomes, it lacks explicit mention of particular stakeholders or specific entrepreneurial outcomes that are highlighted in the REF submission, such as the role of mentorship programs and entrepreneurial ventures, which are impactful yet subtly implied in the AI output.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align closely with the outcomes described in the original REF submission, especially on the technical advancements in gluten-free bread formulation and the mentorship impacts. However, minor aspects, such as specific quantifications

of impact (e.g., turnover) and some industrial applications, were slightly less emphasized.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the outcomes described in the REF submission, especially in recognizing the novel contributions to gluten-free bread technology. Both versions highlight the application of polymer physics principles to address dough instability, referencing the importance of ingredient interactions. However, the AI version slightly underemphasizes specific stakeholder engagements that are pivotal to the REF case.

"
"77. Study Title : Peripheral Impressionisms: challenging perceptions of Impressionism
Rater 1 Rating = 4
AAR Justification: The AI-generated impacts generally align with the key themes of the REF-submitted impacts, including economic, cultural, and academic influences stemming from the exhibitions ""American Impressionism"" and ""Inspiring Impressionism."" The AI-generated impacts correctly recognize how the exhibitions boosted the market for peripheral impressionist artists and raised awareness, contributing to wider public discourse. However, some minor details, such as specific auction results and targeted curator symposia, were more detailed in the REF version.
Rater 2 Rating = 4
AAR Justification: The AI-generated content closely aligns with the REF 2021 submission, accurately reflecting the major outcomes and impacts, especially in highlighting the exhibitions' economic and cultural impacts. For example, the increase in auction prices for Daubigny's works and public engagement through exhibitions were mirrored accurately in both versions.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content closely matches the REF 2021 submission in terms of depicting economic, cultural, and profile-raising impacts, with strong alignment on artist market influence and public engagement. Notable discrepancies include the depth of economic detail in auction results and direct curatorial influence, which are partially summarized but not fully expanded in the AI version.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the REF submission's stated outcomes, particularly in acknowledging the economic and cultural impacts stemming from the exhibitions. AI results also highlight the enhanced profile for lesser-known artists and the economic value increase for Daubigny's works, matching specific achievements noted in REF. Minor differences were observed in the emphasis on secondary cultural impacts on global curatorial practices, which are lightly mentioned in AI-generated content.

"
"78. Study Title : Raising the profile of Scottish Literature through writing and consultancy
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated content captures many of the core impacts in literature and cultural preservation, as seen in Riach's influence on Scottish banknotes and his contributions to tourism through Literary Scotland: A Traveller's Guide. This reflects the REF 2021's emphasis on Riach's role in promoting Scottish literature and culture, though it lacks some specifics around certain consultancy roles.
Rater 2 Rating = 4
AAR Justification: The AI-generated content aligns well with the core objectives described in the REF 2021 submission. For instance, both versions highlight Professor Riach's role in promoting Scottish literature through public channels like RBS banknotes and accessible publications, reflecting the intended impacts on cultural awareness and literacy. However, some minor discrepancies exist in terminology and contextual emphasis. For example, while both versions mention the importance of public engagement, the AI-generated impact places less emphasis on the detailed role of specific literary quotes selected for inclusivity across Gaelic, Scots, and English, which was a significant feature in the original case study.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with the REF submission in key areas, such as the public acclaim for the Royal Bank of Scotland Banknotes, the influence of the Literary Scotland: A Traveller's Guide on tourism, and cultural recreation's role in public engagement. Both versions highlight significant impacts in cultural promotion, heritage awareness, and tourism. However, the AI version lacks some specific examples of public reception and institutional comments featured in the REF submission.
Rater 4 Rating = 4
AAR Justification: The ChatGPT-generated impacts show a high alignment with the original research's objectives and key themes, including the promotion of Scottish literature, public engagement through national newspapers, and contributions to cultural preservation through currency design. The AI-generated impacts correctly emphasize these aspects and echo the

project's cultural and educational goals. Minor discrepancies arise in the specific contributions to public policy, as some AI interpretations were more general than the detailed public engagement activities and outcomes illustrated in the REF submission.

"
"79. Study Title : Secukinumab becomes the first interleukin-17A inhibitor approved for psoriatic arthritis
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts largely align with the original case study, accurately reflecting key aspects of the study's contributions to public health, the economic benefits to the pharmaceutical industry, and the impact on patient outcomes. Both versions highlight secukinumab's global approval, integration into clinical guidelines, and its substantial benefits for patients. The REF submission, however, included specific data on clinical trials, such as percentages of patient improvements, which adds rigor and supports real-world applications.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts are largely accurate in reflecting the fundamental outcomes of the original research. Key impacts—such as the worldwide approval and adoption of secukinumab and the direct benefits to patients through improved quality of life and clinical practice guidelines—are well-aligned with the REF submission. However, there are subtle differences in phrasing that could affect precise alignment. For example, the original report emphasizes McInnes' role in influencing clinical guideline updates through EULAR and GRAPPA, which the AI summary covers in broader terms.
Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts largely align with the REF 2021 submission by capturing key themes, including the therapeutic impact of IL-17A inhibition, clinical benefits of secukinumab, and implications for rheumatology. The alignment is generally strong with some minor variations in detail, particularly regarding specific trial outcomes and regulatory milestones mentioned in the REF submission.
Rater 4 Rating = 5
AAR Justification: The AI-generated impacts and REF 2021 submission align accurately, reflecting the significant contributions of Secukinumab on PsA treatment. Both versions detail the pharmacological background, clinical outcomes, regulatory approvals, and patient benefits. The emphasis on IL-17A as a critical target is consistent in both, highlighting alignment with the original study's clinical impact on patient quality of life, pain management, and disease progression.

"
"80. Study Title : Establishing the Facts, Developing Professionalisation and Enabling Transparent 'Pilgrim-centred' Communication in the UK Hajj Sector
Rater 1 Rating = 4 - Very Good
AAR Justification: The ChatGPT-generated content accurately represents key impacts on policy, professionalization, and stakeholder collaboration, aligning closely with the REF submission. Both versions emphasize the All-Party Parliamentary Group (APPG) on Hajj & Umrah's influence, the role of research in professionalizing UK Hajj services, and its impact on public discourse regarding pilgrim welfare and fraud. ChatGPT accurately included significant references to UK government bodies and regulatory impact.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impact report closely mirrors key areas described in the REF submission, specifically aligning with McLoughlin's work on the regulatory challenges, commercial aspects, and public welfare implications within the Hajj industry. The AI-generated impacts accurately reflect public engagement themes and policy contributions to the UK's Hajj sector, though specific citations, such as instances of media engagement in different countries, were less comprehensive in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: Both the REF submission and ChatGPT output accurately reflect key impacts on governance, fraud prevention, and professionalization in the Hajj sector. ChatGPT captures aspects of the APPG's role and the focus on pilgrim welfare. However, certain nuances about the exact roles of key stakeholders, like Trading Standards or the LHO's significance in governance, are somewhat less explicit in the ChatGPT version.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts are well-aligned with the REF-submitted version, reflecting key aspects such as the focus on policy influence, media engagement, and professionalization of the Hajj sector. Both versions discuss the importance of governance in the Hajj sector and stakeholder engagement, especially with APPG on Hajj & Umrah and CBHUK. However, some nuanced impacts related to specific media outreach and direct APPG interactions

could be better integrated.

"
"81. Study Title : A Bridge for Spies: Overcoming the Practitioner-Academic Gap in Intelligence and Security
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated version aligns well with the primary impacts in the REF submission, capturing key themes of enhancing governmental practices with academic insights and impacting national security, education, and policy. Both versions emphasize intelligence professionalization, cross-departmental standards, and long-term educational frameworks. However, minor misalignments exist in finer policy details and specific agency engagements.
Rater 2 Rating = 4: Very Good
AAR Justification: The AI-generated impacts align well with core outcomes of the REF-submitted version, particularly in identifying improvements in intelligence collaboration protocols and influence on policy-making structures (such as the Intelligence Academy and protocols developed with the NCA and NATO). However, the AI version generalizes some impacts where the REF submission provides specific institutional endorsements (e.g., from the Cabinet Office and NATO), resulting in a minor gap in direct references.
Rater 3 Rating = 4
AAR Justification: Both the ChatGPT-generated impacts and REF-submitted impacts align well, particularly regarding the emphasis on improving engagement between intelligence agencies and academics. For instance, both versions highlight the establishment of professional standards and protocols with agencies like the NCA and NATO, although the REF submission provides more nuanced details of specific agency outcomes.
Rater 4 Rating = 4: Very Good
AAR Justification: The AI-generated impacts closely align with the themes, goals, and identified stakeholders in the REF submission. Both versions emphasize the development and professionalization of intelligence engagement through academic partnerships. However, the REF version provides specific agency examples and exact language regarding protocol impacts, such as the structured interactions fostered within NATO and the NCA, which ground the alignment more concretely.

"
"82. Study Title : Influencing Labour Standards and Stakeholder Action Through International, European and National Law and Policy
Rater 1 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated version aligns closely with REF-submitted content, particularly around the international, European, and national impacts on labor standards, stakeholder engagement, and the contribution of research to legislative and policy changes. It captures the influence on specific sectors, including the ILO and the European Parliament, similar to the original submission. However, some nuances related to specific case studies and particular outcomes at the organizational level (e.g., Greece, India, and South Africa) are slightly generalized in the AI-generated report.
Rater 2 Rating = 4: Very Good
AAR Justification: The ChatGPT-generated impacts align well with the primary outcomes and themes of the REF submission, particularly around labor standards, collective bargaining, and gender pay equality. For example, both versions highlight impacts on international labor policies, especially through ILO collaborations, emphasizing support for inclusive labor standards and social dialogue. However, the AI-generated impacts sometimes lack the depth in illustrating specific organizational impacts, such as the explicit influence on the ILO, OECD, and UN in shaping their strategic frameworks.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align well with the outcomes presented in the REF submission, accurately covering the major influences on labor policy, stakeholder involvement, and collective bargaining. However, minor discrepancies were observed, where AI interpretations emphasized global applicability over specific policy nuances in Greece and the EU.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the original REF submission, particularly in capturing key areas such as international labour policy, gender equality initiatives, and collective bargaining impacts. However, the AI version sometimes lacks specific institutional references and depth regarding unique regional outcomes, such as the European Court of Human Rights integration in policy recommendations.

"
"83. Study Title : Creating Value and Transforming Lives through Arts and Creative Media Practice
Rater 1 Rating = 4

AAR Justification: The AI-generated impact assessment aligns well with the REF-submitted impacts, accurately reflecting core outcomes like heritage preservation, community empowerment, and policy influence. Both versions discuss transformative cultural and social impacts, such as the advocacy for industrial heritage and the participatory urban regeneration efforts in Gdansk and Glasgow, and the significance of community engagement through initiatives in Dharavi. However, the AI-generated impact occasionally generalizes where specific data was presented in the REF submission, leading to slight misalignment in detailed areas.

Rater 2 Rating = 4: Very Good

AAR Justification: The AI-generated impacts align closely with the key elements highlighted in the REF submission, covering community engagement, heritage preservation, and policy influence. It accurately mirrors the focus on industrial heritage (e.g., Gdansk Shipyard) and social change through arts-based methodologies. However, minor discrepancies exist in emphasizing certain geographical impacts.

Rater 3 Rating = 4 (Very Good)

AAR Justification: The ChatGPT-generated version aligns well with the REF-submitted impacts, accurately reflecting areas like urban regeneration, community engagement, and cultural heritage preservation in sites like Gdansk Shipyard and Dharavi. Both versions discuss community empowerment through arts and heritage preservation, though the ChatGPT version simplifies some contextual aspects, such as the local policy impact specifics that are detailed in the REF submission.

Rater 4 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts generally align with the major thematic areas in the REF submission, particularly around community empowerment, cultural heritage preservation, and environmental awareness. However, some specific examples, such as the direct impact on the Gdansk shipyard preservation efforts, are less explicitly captured in the AI version.

"
"84. Study Title : A Transformation in Creep Condition Monitoring for High Temperature, High Pressure Components

Rater 1 Rating = 4 - Very Good

AAR Justification: The AI-generated impacts closely align with the core outcomes in the REF submission, reflecting key impacts across technological, academic, and industrial domains (such as improvements in welding processes, material fatigue life, and structural integrity in high-stress applications). However, slight deviations exist, especially where AI generalizes certain impacts (e.g., societal or environmental impacts) that were detailed with specific examples in the REF version.

Rater 2 Rating = 4 (Very Good)

AAR Justification: The AI-generated impacts closely align with the fundamental advancements and industrial applications emphasized in the REF submission. Specific points such as improvements in monitoring techniques, industrial application in sectors like power generation and chemical processes, and references to technological innovations are accurately represented in the AI version. However, some nuanced details on technical standards (e.g., ASME code submission specifics) were only mentioned in the REF version, affecting full accuracy.

Rater 3 Rating = 4 - Very Good

AAR Justification: The AI-generated impacts closely align with the REF-submitted impacts in terms of the central themes, particularly regarding the advancements in creep condition monitoring, the technological benefits for the industry, and the critical role of residual stress research. Examples that match include the emphasis on WeldCore's role in high-temperature, high-pressure component monitoring and its implications for industries like power generation and chemical processing. However, the AI report was less explicit in detailing certain industrial applications, such as ESKOM's role in load-shedding reduction, which was emphasized in the REF submission.

Rater 4 Rating = 4

AAR Justification: The ChatGPT-generated impacts generally align with the outcomes detailed in the REF submission, covering the primary research achievements in high-temperature condition monitoring, industrial application in ESKOM, and technological adoption at SASOL. However, while both versions mention core innovations like WeldCore technology, the REF submission provides more detail on the specific industry and academic partnerships that underpinned these advancements.

"
"85. Study Title : The creation (and re-creation) of contemporary female heroines at the center of new plays for the theatre.

Rater 1 Rating = 4 - Very Good

AAR Justification: The AI-generated impacts accurately reflect the original project's emphasis on creating substantial roles for women in theatre, enhancing international representation, and

catalyzing public debates on gender in theatre. For instance, both versions capture Harris's role in changing programming practices and boosting representation for female-driven narratives on major stages (e.g., Royal Court Theatre and Edinburgh International Festival). However, there are slight deviations regarding specific cultural impacts and awards.
Rater 2 Rating = 4
AAR Justification: The AI-generated impacts align well with the REF submission, particularly capturing the project's focus on elevating female representation in theater. Specifics, such as the international reach and the challenge of traditional female roles, were echoed in both versions. However, certain nuanced aspects, like the mention of cultural contexts (e.g., Turkish theatre's response) were more detailed in the REF submission.
Rater 3 Rating = 4
AAR Justification: The AI-generated content aligns closely with the REF 2021-submitted version by recognizing the impact of Zinnie Harris's plays on gender representation in theatre, the public debate on women's roles, and the influence on theatre programming. However, it slightly lacks in-depth detail on specific outcomes, such as the exact reach of each play across various regions and exact attendance figures, which were meticulously outlined in the REF submission.
Rater 4 Rating = [4 - Very Good]
AAR Justification: The AI-generated impacts align well with the central themes of Harris's work, particularly in placing women at the forefront of narrative development. While the AI does recognize Harris's focus on female protagonists and representation in theater, it lacks specificity on some theatre production metrics and precise geographical influence outlined in the REF submission.

"
"86. Study Title : Changes to cervical screening policies following the rollout of human papillomavirus (HPV) vaccination
Rater 1 Rating = 4 - Very Good
AAR Justification: Both versions emphasize the significant health, policy, and societal impacts of HPV vaccination research. The REF submission provides specific references to changes in the Scottish Cervical Screening Programme and UK policy recommendations, which are addressed in the ChatGPT-generated version, albeit with broader mentions of "policy influence."
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with key outcomes in the original REF case, particularly the public health benefits of reduced HPV prevalence, policy changes in the Scottish Cervical Screening Programme, and the influence on UK and international vaccination policy. Specifics, such as the shift in cervical screening age, are captured, yet nuanced stakeholder perspectives (e.g., feedback from Health Protection Scotland) appear less explicitly in the AI version.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts largely align with the original case study outcomes, accurately reflecting the primary contributions to public health policy and awareness. Both versions mention the influence on cervical screening practices, vaccination advocacy, and global health policies.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts accurately reflect the project's core outcomes, such as policy influence on cervical screening and international public health policy. Notably, it captures key influences like adjustments to age ranges in screening programs, reductions in cervical disease prevalence, and global awareness of HPV vaccine success. However, some nuances, such as the specific influence on UK cervical screening improvements, are less detailed.

"
"87. Study Title : Combating Crop Losses and Improving Global Food Supplies through Mathematical Modelling of "Gene Silencing"
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align closely with the REF 2021 submission, accurately reflecting the project's focus on sustainable agricultural practices, crop yield improvements, and environmental benefits of microbial biostimulants. Specific outcomes, like yield increases across various crops and reduced dependency on pesticides, are well captured in the AI version, though some nuanced points, like certain regional impacts and exact figures for yield increases, were better detailed in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the REF-submitted impacts, capturing key areas such as increased crop yields, cost reductions, and the environmental benefits of RNAi-based microbial biostimulants. The ChatGPT version accurately reflects core outcomes, including the significant role of the project in reducing reliance on pesticides and promoting sustainable agricultural practices in Ukraine. However, some specific stakeholder

perspectives, like direct farmer testimonies, were less prominent in the AI version, affecting the completeness of alignment with specific evidence in the REF.
Rater 3 Rating = 4
AAR Justification: The AI-generated content aligns with the primary impacts in the REF submission, accurately reflecting agricultural, economic, and environmental impacts. Key elements such as pest management and yield improvements were presented, matching the REF document. However, some nuanced details, such as regional economic benefits, were less explicit in the AI version.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts closely reflect the main objectives and outcomes of the original research. The REF-submitted study emphasizes RNAi's potential to control crop pathogens and the AI-generated version successfully identifies and describes the relevance of RNAi applications in agriculture. However, some detailed metrics (e.g., specific crop yield improvements and region-specific effects) noted in the REF submission are generalized in the AI version.

"
"88. Study Title : Emerging Media, Learning, and Organisational Practice - Driving Change in Tourism and Education in Northern Ireland
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the REF-submitted impacts, accurately reflecting the project's emphasis on enhancing tourism and education through emerging technologies such as augmented reality (AR) and immersive learning environments. Both versions discuss the technological advancements in tourism and educational impacts, although the REF version provides more precise detail on specific projects and stakeholders involved, such as local councils and educational bodies. The AI version slightly generalizes some points but retains core themes accurately.
Rater 2 Rating = 4 - Very Good
AAR Justification: The AI-generated content demonstrates a high degree of alignment with the REF-submitted impacts by accurately reflecting the outcomes related to tourism sector advancement, technological adaptation, and educational enhancements. For instance, it accurately captures Jackson's influence on integrating mobile app solutions in tourism and the socio-cultural impact on Northern Ireland's tourism strategies. However, the AI version occasionally generalizes some specific stakeholder feedback (e.g., from CCGBC's managers and MHT representatives), lacking the personal perspective evident in the REF submission, which slightly affects the alignment.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts closely align with the core outcomes outlined in the REF submission, especially in aspects like tourism enhancement and educational impacts. Key elements such as the integration of immersive technologies and data-driven tourism are well-reflected. However, some subtle aspects, like long-term impacts on policy frameworks and detailed insights into regional applications, are less emphasized in the ChatGPT version.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the core outcomes of the REF 2021 submission, capturing key aspects of technological innovation in tourism and education. Both sources emphasize the influence of AR and immersive technologies in enhancing the tourist experience and pedagogical practices, aligning with the project's objectives. However, while AI mentioned the general impacts on educational practices and tourism, it missed nuances like Jackson's specific contributions to the ""virtual journey"" concept and precise details on user experience modifications in the tourism sector, slightly lowering its alignment score.

"
"89. Study Title : Global adoption of the Dermatology Life Quality Index (DLQI) into clinical practice
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts are highly aligned with the original REF submission, accurately reflecting the significance of DLQI in dermatology, particularly its role in assessing quality of life and influencing clinical guidelines. Both versions highlight DLQI's widespread adoption, use in clinical trials, and its pivotal role in clinical practice across multiple countries. However, some nuanced aspects of clinical validation and real-world trial applications are emphasized more in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: Both versions address the central impact of DLQI on dermatology, accurately aligning with real-world clinical, pharmaceutical, and policy applications. However, the ChatGPT-generated impact missed some detailed aspects, such as the NICE Technology Appraisals for specific conditions and specific revenue figures from licensing in pharmaceutical contexts,

which were highlighted in the REF submission.
Rater 3 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align well with the original case study, accurately reflecting core elements such as DLQI's global clinical adoption, influence on healthcare policies, and economic impacts. However, minor discrepancies exist in the finer contextual details. For example, the AI version effectively covers DLQI's use in clinical guidelines and pharmaceutical trials but lacks specific references to NICE approvals and exact revenue figures, which are included in the REF submission.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts are largely accurate and reflect the primary outcomes discussed in the REF 2021 submission, such as the adoption of DLQI in clinical trials and its inclusion in guidelines. However, some nuanced details on specific guideline developments and economic revenue specifics are less pronounced in the AI version. The AI content successfully mirrors the broader adoption impact but could improve in precisely matching the language on clinical guideline inclusions.

"
"90. Study Title : Improving the healthcare experiences of children and young people
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts are closely aligned with the primary objectives and impacts described in the REF 2021 submission, specifically in areas such as improving clinical practice through child-centered approaches, influencing policies for pediatric care, and enhancing educational frameworks for healthcare professionals. Notably, both versions emphasize the emotional and psychological needs of children undergoing medical procedures, with explicit mentions of reducing procedural distress through interventions like the "clinical pause" and multimedia resources, such as the 'Children Coming to Hospital' digital platform. However, some AI-generated content generalized aspects, for example, extending certain emotional well-being impacts to a broader pediatric population, rather than the specific cognitive and developmental needs described in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with key outcomes highlighted in the REF 2021 submission, such as improvements in healthcare practices, policy influence, and health literacy for children. For example, both versions emphasize the impact on clinical practices related to physical restraint and holding, reflecting this aspect accurately. However, AI content occasionally lacks precise citations or contextual detail.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts align closely with the core impacts in the REF submission, covering aspects like improving procedural health literacy, informing clinical holding practices, and enhancing understanding of pain management in children. However, some specific policy influences and detailed impacts on stakeholder engagement, such as involvement with the Restraint Reduction Network (RRN) in the REF submission, are less emphasized in the AI-generated version.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts align well with the original REF submission, reflecting the research's emphasis on pediatric pain management, caregiver education, and support in handling non-verbal children with cognitive impairments. Both versions emphasize practical outcomes and actionable recommendations for health providers and caregivers, highlighting effective communication and training needs. This shows a strong alignment with the REF's goals.

"
"91. Study Title : The value of the Carers' Alert Thermometer (CAT) in identifying family carers' needs and supporting them in their caring role
Rater 1 Rating = 4
AAR Justification: Both versions consistently highlight CAT's effectiveness in assessing carers' needs, its role in reducing carer strain, and its contribution to policy and healthcare practices. The REF version includes specific endorsements from institutions like NIHR and detailed feedback from carers, such as expressions of feeling validated, which deepen the accuracy and grounding in real-world contexts. ChatGPT addresses CAT's foundational role in healthcare improvements for carers but generalizes some details about institutional support and specific stakeholder engagement.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the core outcomes detailed in the REF submission, such as the CAT's role in assessing and prioritizing carers' needs and its adoption across various health care settings nationally and internationally. However, minor discrepancies exist in reflecting the specific stakeholder engagements and some unique policy

recommendations explicitly mentioned in the REF version, which slightly detract from a perfect alignment.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impacts accurately reflect the main impacts in health, social care, and policy reform as described in the REF submission. The ChatGPT report emphasizes CAT's contributions to carer well-being, policy influence, and international adoption, which align closely with the core impacts cited in the REF case. Both versions highlight the CAT's significance in carer assessment and the subsequent adoption by palliative care and community health organizations, showing strong alignment in key impacts.
Rater 4 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align closely with the REF submission in identifying CAT's role in caregiver support, policy influence, and adoption in health and social care. The AI report includes most core areas such as policy influence and international adoption but lacks some specificity in examples.

"
"92. Study Title : God-complexity and the Multiple God-Aspects Framework
Rater 1 Rating = 3 - Good
AAR Justification: The AI-generated impacts generally align with the REF 2021 submission by addressing core themes such as psychological and theological dimensions of God representations and their implications on mental health, spirituality, and well-being. However, nuances, like the detailed application in specific populations or direct quotes from stakeholders, present in the REF submission are less emphasized in the AI output, creating minor discrepancies in accuracy.
Rater 2 Rating = 4
AAR Justification: The AI-generated impacts closely align with the REF 2021 case study on foundational aspects, particularly regarding the aim to empirically assess the complexity of God-representations and the applications in clinical practice and academic discourse. Both versions emphasize the empirical foundation and psychological aspects of understanding God-complexity, addressing intersections of belief and identity (e.g., social cognition and self-complexity), which is central to the REF submission. The AI version, however, does not fully match the REF submission's emphasis on the public-facing communication efforts and specific presentation contexts within conferences.
Rater 3 Rating = 4
AAR Justification: Both the AI-generated impact and REF submission align closely in presenting the foundational goals and outcomes of the research, such as expanding the understanding of God representations and developing the God-complexity measure. However, the REF submission provides additional context on specific stakeholder engagement in religious communities and details feedback received from academic presentations.
Rater 4 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the foundational aspects of the project, including its focus on God-complexity and related psychological frameworks, reflecting a strong understanding of its objectives in influencing both academic and community perceptions. However, there are minor deviations in contextual emphasis, particularly in aspects such as specific community engagement outcomes cited in the REF submission but less highlighted in the AI version.

"
"93. Study Title : Abertay Game Lab: play, performance, and public engagement with games
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the detailed impacts outlined in the REF 2021 submission. The AI effectively captures the project's influence on cultural institutions, the arts, and public engagement, particularly regarding the V&A museum's exhibitions and collaborations, and the broader contributions to public play and video game curation. The AI version also identifies impacts on the games industry and cultural organizations.
Rater 2 Rating = 4
AAR Justification: The AI-generated impacts largely align with the REF-submitted version, capturing major themes such as the influence on arts practitioners, cultural organizations, and the games industry. The AI analysis reflects the broad impacts on stakeholders, including those involved in curating video games and enhancing public engagement through exhibitions and virtual reality experiences. For example, the AI mentions impacts related to social play and game curation, aligning well with the original REF details, such as the V&A museum exhibition and its visitor demographics. However, some nuanced relationships and specific impacts, like V&A visitor demographics (first-time visitor rates, age distributions), were only lightly addressed in the AI-generated version.

Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns closely with the primary outcomes and impacts, specifically in domains like public engagement and the creative arts. Specific details regarding game curation and audience development, however, were better emphasized in the REF submission, while ChatGPT provided general alignment but missed nuances like demographic impacts on museum audiences.
Rater 4 Rating = 4
AAR Justification: The ChatGPT-generated impacts generally align well with the REF 2021 submission, reflecting key themes of game curation, social engagement, and interdisciplinary impact within gaming and arts. Specific impacts on public policy, such as influences on museum practices and cultural initiatives, were mirrored with accuracy. However, some nuanced details on the regional focus (e.g., specific workshops in Dundee and London) were only briefly referenced, affecting the rating.

"
"94. Study Title : Improving Care for Patients with Chronic and Distressing Tinnitus through Mindfulness Based Interventions
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impacts align closely with the REF submission by capturing significant areas such as patient benefits, public awareness, and clinical adoption. It accurately reflects the emphasis on mindfulness for tinnitus, supported by detailed outcomes related to patient satisfaction and clinical guidelines. However, some nuances, such as specific patient testimonies and feedback from major clinical centers, are less emphasized in the AI version.
Rater 2 Rating = 4
AAR Justification: The AI-generated impacts align well with the central objectives and achievements outlined in the REF 2021 submission, accurately reflecting the use of MBCT-t in transforming tinnitus care through clinical practice changes and the adoption of mindfulness interventions. Specific policy influences, like integration into NICE guidelines, are captured, showing a strong alignment. However, some nuanced details regarding the extent of public awareness initiatives and the professional training impact are not as emphasized in the AI-generated text.
Rater 3 Rating = 4 (Very Good)
AAR Justification: Both the ChatGPT and REF-submitted versions largely align in representing the scope and impact of MBCT on tinnitus treatment, including benefits to clinical practice and patient satisfaction. The core aspects of clinical guidelines and policy adoption are reflected in both.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts are generally aligned with the specific outcomes noted in the REF submission, especially regarding clinical adoption, public awareness, and the economic impact of tinnitus treatment with MBCT-t. However, some nuanced aspects such as specific endorsements by individual hospitals and precise feedback on patient satisfaction rates were not fully captured. This makes the alignment robust but with some gaps in detailed accuracy.

"
"95. Study Title : Developing and Embedding Effective Careers Guidance for Young People in England
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align closely with the themes in the REF submission, accurately reflecting the project's contributions to policy (e.g., the adoption of Gatsby Benchmarks in career guidance policy) and educational improvement in career readiness. Both versions emphasize the social and educational benefits for students, the integration of career guidance in schools, and the role of benchmarks. However, some direct influences, such as the high-level policy discussions or direct mentions of implementation in specific regions like Hong Kong or Wales, are detailed more in the REF submission.
Rater 2 Rating = 4: Very Good
AAR Justification: The ChatGPT-generated impacts align well with the original study's outcomes on multiple fronts, such as the development and adoption of the Gatsby Benchmarks and its influence on educational policies in England and beyond. Key impacts, such as the influence on public policy through the Department for Education's Careers Strategy and the practical implementation of Career Leaders, are accurately reflected. However, certain nuanced details on international policy influences in specific regions like Hong Kong and Wales, though broadly mentioned, are less precisely articulated in the AI-generated impacts.
Rater 3 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts closely align with the REF-submitted impacts. Both versions emphasize the Gatsby Benchmarks' influence on career guidance in English schools and

its integration into the Department for Education Careers Strategy. However, some specific details are missing in the AI version, such as Baroness Bottomley's endorsement and specific reference to the House of Lords debate, affecting complete alignment.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts closely align with the key themes presented in the REF-submitted version, notably on the educational, policy, and social domains. However, minor discrepancies were noted in how some aspects, such as implementation nuances and specific international adaptations, were conveyed. While the AI captures the core principles of the Gatsby Benchmarks' application in schools and colleges, it underrepresents certain geographic distinctions and stakeholder engagements specified in the REF submission.

"
"96. Study Title : Design meets disability: changing the relationship between disability and design in business, culture, practice, and education
Rater 1 Rating = 4 - Very Good
AAR Justification: The ChatGPT-generated impacts closely align with the original REF case study, matching key themes in disability design and inclusive practices. Both versions emphasize major impacts on public awareness and cultural perceptions (e.g., The Valuable 500 initiative in business and Alleles in personal empowerment). Although highly aligned, minor nuances in the industry engagement strategies are more clearly detailed in the REF submission.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The AI-generated impacts align well with the core themes and objectives of the REF submission, particularly in addressing the empowerment of individuals with disabilities, promoting inclusivity, and challenging traditional design paradigms. However, certain specific events and stakeholder quotes, critical to the narrative and impact context in the REF, are either generalized or absent in the AI version, slightly impacting overall accuracy. Examples of alignment include both versions' focus on Alleles' influence on self-expression and The Valuable 500's policy impact in corporate inclusivity.
Rater 3 Rating = 4 (Very Good)
AAR Justification: Both the REF submission and ChatGPT version align on the overarching goals, such as promoting inclusivity in design and shifting societal perspectives on disability. Both mention Alleles and The Valuable 500 as case studies reflecting impactful societal changes in framing disability. However, ChatGPT version slightly generalized some impacts, which affected precision in industry-specific details.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts closely align with the original study's focus on reframing disability through design in multiple business and cultural contexts. The main aspects, such as Alleles' contributions to self-expression for prosthetic users and The Valuable 500's advocacy for inclusion at the executive level, were well-covered. Some nuances of cultural influence were less emphasized.

"
"97. Study Title : Digital Archiving for Curation and Dissemination
Rater 1 Rating = 4
AAR Justification: Both versions effectively capture the core impact of preserving and democratizing access to historical and cultural assets, such as the Demarco Archive and The Water Hen exhibition. However, the ChatGPT version lacked nuanced examples of specific stakeholders' contributions, such as the UNESCO and the Tate's acknowledgment. This results in a slight misalignment in specificity compared to the REF version's emphasis on international curation recognition (e.g., Venice Biennale).
Rater 2 Rating = 4
AAR Justification: AI accurately covered cultural and academic impacts but missed specific public policy impact discussed in REF.
Rater 3 Rating = 4
AAR Justification: The AI-generated content aligns well with the REF submission, reflecting the project's impacts on cultural preservation, academic engagement, and public outreach. Specific references to archival digitization and impact on exhibition practices match those detailed in the REF submission, especially in enhancing access to Eastern European cultural assets.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts show substantial alignment with the outcomes and objectives from the REF 2021 submission. Both versions address cultural preservation, with detailed mentions of the Demarco Archive, particularly emphasizing digital preservation efforts that parallel REF's core objective of conserving cultural artifacts. However, the REF 2021 version provides additional nuanced details about specific stakeholder feedback, like citations from scholars and exhibition statistics, which were not as deeply highlighted in the AI-generated version.

"

"98. Study Title : Prevention and management of head injuries in cricket and rugby union
Rater 1 Rating = 4 - Very Good
AAR Justification: The ChatGPT version aligns closely with the REF submission by covering injury surveillance, helmet safety, and concussion education in cricket and rugby, including details on helmet design changes and sports policy influences (e.g., the ICC helmet standards). Although generally well-aligned, a few specific outcomes in the REF case study, such as longitudinal funding influences and exact stakeholder feedback mechanisms, were less detailed in the AI version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: The ChatGPT-generated impact largely aligns with the REF 2021 submission in identifying injury risks and advancements in player safety, particularly for helmet and concussion protocols in cricket and rugby. Examples include acknowledgment of helmet safety standards and the surveillance methods to reduce concussions, which are central in both versions.
Rater 3 Rating = 4
AAR Justification: The AI-generated impacts align closely with the outcomes of the REF 2021 submission, particularly in terms of injury prevention and safety standardization efforts. Both versions detail significant achievements in helmet safety standards in cricket and the Concussion Education Programme in rugby. The AI's additional insights into long-term health outcomes also support the project's objectives.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts largely align with the REF 2021 submission, accurately reflecting the project's main outcomes, such as helmet safety in cricket and the concussion education program in rugby. Key safety standards and policy changes are well represented, and there is sufficient recognition of the stakeholders (e.g., ICC, WRU, ECB). However, some nuanced elements, like specific injury definitions used in surveillance, were underemphasized.

"

"99. Study Title : The Kindertransport 1938/1939 to the UK: History Informing the Future
Rater 1 Rating = 4 - Very Good
AAR Justification: The AI-generated impact aligns well with the overall goals and findings of the REF-submitted content, particularly in addressing the key areas of social impact, policy implications, and educational influence that the original case study highlights. Both versions emphasize the public awareness raised by the Kindertransport research, its educational contributions, and its effect on understanding and supporting contemporary refugee policies. The AI-generated version also captures the historical and policy-related nuances presented in the original submission, though it occasionally omits specific program names and precise stakeholder feedback included in the REF 2021 version.
Rater 2 Rating = 4 (Very Good)
AAR Justification: Both versions align closely on key themes such as historical trauma, public awareness, and policy impact related to the Kindertransport, accurately reflecting the historical and policy implications for contemporary refugee issues. Both discuss the educational and societal benefits of exhibitions and the impact on public discourse, though the AI version is more generalized in places.
Rater 3 Rating = 4 - Very Good
AAR Justification: Both versions accurately reflect the core outcomes of the Kindertransport project, covering its impact on public understanding, policy influence, and educational outreach. The ChatGPT version aligns well with most elements, accurately capturing public awareness initiatives and their effects on refugee policies. However, it lacks certain nuanced reflections on specific public talks and exhibitions noted in the REF submission.
Rater 4 Rating = 4
AAR Justification: The AI-generated impacts largely align with the main themes and outcomes of the REF submission, focusing on historical awareness, public policy implications, and social engagement, as shown in both cases. However, the AI version lacks a few nuanced details regarding the specific stakeholders and long-term impact on social perceptions that the REF submission covers.

"

"100. Study Title : An evidence-based approach reduces the local costs of biodiversity conservation in low- and middle-income countries
Rater 1 Rating = 4 (Very Good)
AAR Justification: The AI-generated content aligns well with most key outcomes described in the REF submission, such as the influence on policy reform in Madagascar and internationally (e.g., use of research findings in biodiversity offset practices). Both versions highlight the role of

biodiversity offsets in addressing local costs and the impact of REDD+ initiatives on local communities in Madagascar. However, minor discrepancies were noted regarding some specific impact claims, leading to a rating of 4 rather than 5.

Rater 2 Rating = 4: Very Good

AAR Justification: The AI-generated impacts align closely with the original REF case study's emphasis on biodiversity conservation's social, environmental, and policy effects, particularly on community livelihoods and policy reform efforts in Madagascar. However, some specific nuances regarding multi-national stakeholder engagement (e.g., Ambatovy mine case) and the chronological sequence of influence on policy adaptations were more detailed in the REF submission.

Rater 3 Rating = 4 (Very Good)

AAR Justification: The ChatGPT version aligns closely with the REF submission, covering most of the specific impacts mentioned, especially regarding biodiversity offsets and the social-economic impacts on local communities. The generated version accurately captures the study's influence on policy, corporate accountability, and community engagement, aligning well with the original's emphasis on Madagascar's biodiversity projects.

Rater 4 Rating = 4 (Very Good)

AAR Justification: ChatGPT-generated content largely aligns with the detailed outcomes presented in the REF submission, capturing the influence on public policy, industrial practices, and social equity within biodiversity offsets. ChatGPT accurately identified the policy adjustments in Madagascar and its role in informing practices for other major conservation stakeholders like USAID. It also highlights how Bangor's principles influenced project decisions in Myanmar and Mozambique.

"